이학석사학위논문

Single-molecule Fluorescence Study on Structural Roles of Guide RNAs in Activation of Cas9 Endonuclease

Cas9 단백질 활성화를 위한 가이드 RNA 의 구조적 역할에 대한 단분자 형광 연구

2017년 2월

서울대학교 대학원

화학부 물리화학전공

성 기 원

MS. Dissertation

Single-molecule Fluorescence Study on Structural Roles of Guide RNAs in Activation of Cas9 Endonuclease

February 2017

Research Advisor:

Professor Seong Keun Kim

Department of Chemistry

Seoul National University

Keewon Sung

Abstract

Single-molecule Fluorescence Study on Structural Roles of Guide RNAs in Activation of Cas9 Endonuclease

Keewon Sung

Major in Physical Chemistry

Department of Chemistry

The Graduate School

Seoul National University

Clustered regularly interspaced short palindromic repeats (CRISPR) / CRISPR-associated (Cas) system has recently been discovered as a prokaryotic adaptive immune mechanism against invading viruses and plasmids. Among various types of the CRISPR/Cas systems, type-II consists of Cas9 endonuclease and two guide RNAs (gRNA), CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA), whose ribonucleoprotein complex recognizes and cleaves target DNA. Although extensive structural studies of Cas9-gRNA complex has been carried out since the CRISPR/Cas9 system was successfully employed to genome editing, the detailed mechanism and dynamics of gRNAs in the structural activation of the Cas9 endonuclease are still elusive.

Herein, we review our studies on the structural roles of the two guide RNAs in the activation of Cas9 endonuclease using pseudo-ensemble and single-molecule fluorescence spectroscopic assays, thereby providing mechanistic details which are not available from other static and/or ensemble-averaged measurements. Firstly, tracrRNA was found to be critical in the formation of functional Cas9-gRNA complex (Cas9:gRNA). In the absence of tracrRNA, the Cas9 protein became inactivated so that it failed to form Cas9:gRNA, whereas tracrRNA-Cas9 interaction prevented the inactivation pathway by leading Cas9 toward a complexation pathway for Cas9:gRNA. Secondly, crRNA was found to regulate the nuclease activation of Cas9 after Cas9:gRNA bound to the target DNA with forming Cas9-gRNA-DNA ternary complex (Cas9:gRNA:DNA). R-loop structure between crRNA and the target DNA in Cas9:gRNA:DNA exhibited intrinsically repetitive transitions between two distinct sub-conformations termed ‘open’ and ‘zipped’ conformations. This crRNA-related conformational dynamics was proved to be a crucial factor for the nuclease activation by the observation that the bound DNA was cleaved only when the zipped conformation formed. Overall, these results show mechanistic insights into the gRNA-associated structural activation of the Cas9 endonuclease.

Keywords: Single-molecule spectroscopy, Total internal reflection fluorescence microscopy, Fluorescence resonance energy transfer, CRISPR/Cas9, Structural dynamics, Enzymatic mechanism

***Student Number*: 2015-20386**

Table of Contents

1. Introduction

Recently, prokaryotes (most archaea and nearly 40% of bacteria) are found to have an adaptive immune system, termed clustered regularly interspaced short palindromic repeats (CRISPR) / CRISPR-associated (Cas) system, which defends themselves against foreign nucleic acids (DNA or RNA) from invasive bacteriophages and plasmids[1–3]. The brief mechanism of action of CRISPR/Cas system is as follows: 1) CRISPR RNA (crRNA), which contains individual short sequences ("spacer") obtained from the foreign nucleic acid fragments ("protospacer"), is transcribed from the CRISPR locus of the genome[4], 2) the crRNA, in the form of crRNA-Cas protein effector complex, recognizes the target nucleic acid sequences via Watson-Crick base-pairing with Cas proteins inducing targeted double-strand breaks to destroy the invading nucleic acids[5].

Among the three major types of CRISPR/Cas systems, the type-II CRSPR/Cas system has been exclusively employed to *in vitro* and *in vivo* genome editing due to its simple handling as well as high specificity[6]. In the type-II system, only one protein, Cas9, with two guide RNAs (gRNAs), crRNA and trans-activating crRNA (tracrRNA), is required to form a functional RNA-guided endonuclease complex[7]. The Cas9-gRNA complex (Cas9:gRNA) recognizes and cleaves the target DNA sequences which contain protospacer and protospacer-adjacent motif (PAM) right next to the protospacer sequence[8–10].

Since Cas9 from Streptococcus pyogenes was first applied to genome editing successfully in human cells[11–14], extensive structural studies for the Cas9 system have been reported. The Cas9-gRNA interaction leads Cas9 to Cas9:gRNA, in which Cas9 is structurally distinct from *apo*-Cas9, followed by further transitioning into the

structure of the Cas9-gRNA-DNA ternary complex (Cas9:gRNA:DNA) upon binding to the target DNA, revealed by biochemical assays[15], X-ray crystallography[16–20], and single-particle cryo-electron microscopy[19,20]. These previous reports show that gRNAs are not only involved in R-loop expansion between the target DNA and crRNA but also critical for the protein domain rearrangement of Cas9 for the target DNA recognition and cleavage. However, the details, especially in the aspect of structural dynamics, of the two gRNAs in structural activation of the Cas9 nuclease remain to be elucidated. Therefore, we set out to explore the structural roles of gRNAs in the functional activation of Cas9 at single-molecule level.

This thesis summarizes our single-molecule and single-molecule-based pseudo-ensemble investigations into the molecular mechanism of the two gRNAs on Cas9:gRNA complexation and the conformational changes related to the nuclease activity. In Chapter 2, we discuss the basic principles and major characteristics of the single-molecule fluorescence method and the CRISPR/Cas system, employed in these studies. The main results follow in Chapter 3 and 4; each chapter illustrates the newly unraveled structural roles of tracrRNA and crRNA, respectively. Using the pseudo-ensemble fluorescence spectroscopy, we discovered an inactivated *apo*-Cas9 state, transition toward which was repressed by tracrRNA-Cas9 interaction (Chapter 3). Moreover, real-time observation of Cas9:gRNA:DNA using the single-molecule fluorescence spectroscopy revealed sub-conformational heterogeneity and flexibility in the R-loop structure between crRNA and the target DNA, distribution of which regulated the nuclease activity of the Cas9 nuclease (Chapter 4).

The employment of the single-molecule spectroscopic technique in monitoring Cas9:gRNA functioning results in that tracrRNA plays an initiative role in the formation pathway of functional Cas9:gRNA, whereas crRNA plays a regulatory role in the nuclease activation of Cas9:gRNA:DNA. By providing the mechanistic details on the gRNA-induced structural activation of the Cas9 nuclease, these studies contribute to the thorough understanding of the mechanism of action of the CRISPR/Cas9 system, thereby enabling more efficient genome editing.

2. Basic Principles

2.1. Single-molecule Fluorescence Spectroscopy

2.1.1. Introduction: advantages of single-molecule spectroscopy

Due to stochastic property of the nature, all chemical and thus biological processes occur stochastically, leading to heterogeneities for the chemical or biological complex systems. Most experimental techniques, however, measure ensemble-averaged signals from a large number of molecules at heterogeneous states on the way of the processes, which lacks information on the identification, distribution, and dynamics of the sub-populations. Single-molecule technique overcomes this limitation completely. By detecting molecules one by one, the sub-populations can be readily identified and their distribution also be calculated directly. Furthermore, real-time single-molecule observation enables monitoring the dynamics of transitions between the sub-populations. Owing to these exclusive advantages, the single-molecule spectroscopy has been actively applied to study microscopic characteristics of numerous complex processes, especially in biological systems.

2.1.2. Fluorescence resonance energy transfer (FRET)

Fluorescence (Förster) resonance energy transfer (FRET) is a physically well-established energy transfer process between two objects (FRET pair: donor and acceptor), in which energy from the excited donor directly transfer to the acceptor when the energy levels of the donor and acceptor are resonant. For a pair of fluorophores satisfying the resonance condition, the occurrence of FRET could be readily detected by the acceptor fluorescence with donor excitation. Since FRET

efficiency depends sharply on the inter-fluorophore distance in the range of 2-8 nm, which is consistent with the size of bio-macromolecules, it becomes one of the most powerful fluorescence spectroscopic technique to monitor biological processes involving binding or conformational rearrangement of bio-macromolecules (i.e. proteins and nucleic acids) at the molecular level. Further, combined with single-molecule techniques, single-molecule FRET (smFRET) technique enables detection of transient intermediates, structural heterogeneity and dynamics during the reaction pathways of biological processes.

The FRET phenomenon naturally emerges from the quantum mechanical description of (induced) dipole-dipole interaction between the FRET pair. In the presence of the dipole interaction, the rate of the transition from a collective state of the donor excited and acceptor at the ground toward that of the donor at the ground and acceptor excited, which refers to the rate of FRET, can be expressed in terms of the transition matrix element between the excited state of donor and acceptor according to the Fermi's golden rule:

$$k_{FRET} \propto \left|\left\langle D^*, A\middle|\widehat{H'}\middle|D, A^*\right\rangle\right|^2$$

where D and A represent the ground states with D^* and A^* representing the excited states of the donor and acceptor, respectively, and $\widehat{H'}$ is the interaction Hamiltonian for the dipole-dipole interaction potential. The interaction Hamiltonian, $\widehat{H'}$, can be expressed as:

$$\widehat{H'} = \frac{\overrightarrow{\mu_D} \cdot \overrightarrow{\mu_A}}{R^3} - \frac{3\left(\overrightarrow{\mu_D} \cdot \vec{R}\right)\left(\overrightarrow{\mu_A} \cdot \vec{R}\right)}{R^5}$$

where $\overrightarrow{\mu_D}$ and $\overrightarrow{\mu_A}$ are dipole moments of the donor and acceptor, respectively, and

$\vec{R}$ is the distance vector from the center position of the donor to that of acceptor. The above relation shows that FRET is the non-radiative energy transfer, in which energy from the excited state of the donor (D^*) directly transfers to the ground state of the acceptor (A), thereby the donor quenched to the ground state (D) with the coincident excitation of the acceptor (A^*), followed by the fluorescence emission of the acceptor (Fig. 2.1.).

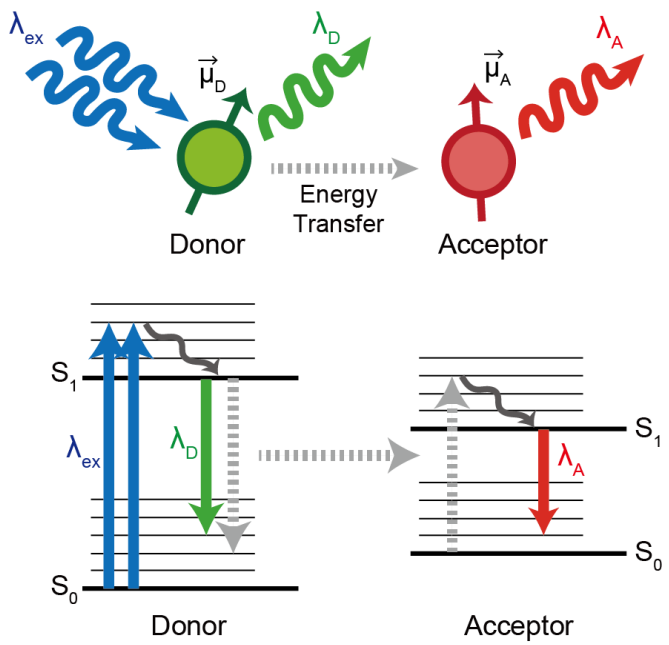

Figure 2.1. Physical principle of FRET with Jablonski diagram. FRET occurs via dipole-dipole interaction of donor ($\overrightarrow{\boldsymbol{\mu_D}}$) and acceptor ($\overrightarrow{\boldsymbol{\mu_A}}$) molecules when the donor emission and acceptor absorption energies are resonant.

An explicit formula for the rate of the energy transfer was firstly derived by Theodor Förster[21]:

$$k_{FRET} = \frac{1}{\tau_D{}^\circ}\left[\frac{R_0}{R}\right]^6 \quad \cdots (*)$$

where R is the inter-fluorophore distance, $\tau_D°$ is the lifetime of the excited state of donor in the absence of FRET (i.e. $\tau_D° = (k^D{}_{fl} + \sum k^D{}_{nr})^{-1}$ where $k^D{}_{fl}$ and $k^D{}_{nr}$ are the fluorescence and any other non-radiative decay rates of the donor, respectively), and R_0 is the characteristic distance, called Förster radius, which depends on the photophysical properties of the two fluorophores of the FRET pair. This equation indicates not only that the FRET rate depends on the inverse sixth power of inter-fluorophore distance but also that the FRET occurs in the time order of $\tau_D°$, which is typically in ns order for excited singlet state. In detail, the Förster radius depends on four other terms according to the following relation:

$$R_0 = (8.79 \times 10^{-5} \cdot n^{-4} \cdot \phi_D \cdot \kappa^2 \cdot J)^{1/6} \text{ [Å]}$$

where n is the refractive index of the medium, ϕ_D is the fluorescence quantum yield of the donor, κ^2 is the factor for relative orientation between the dipole moments of the donor and acceptor, and J in the unit of [$nm^4M^{-1}cm^{-1}$] is the spectral overlap between donor emission and acceptor absorption spectra. For the orientation factor, κ^2, it can be calculated by the following equation (Fig. 2.2, left):

$$\kappa^2 = (\cos\theta_{DA} - 3\cos\theta_{DR}\cos\theta_{AR})^2$$

where θ_{DA} is the angle between the donor dipole moment, $\overrightarrow{\mu_D}$, and acceptor dipole moment, $\overrightarrow{\mu_A}$, θ_{DR} is the angle between $\overrightarrow{\mu_D}$ and the inter-fluorophore distance vector, $\vec{R}$, and θ_{AR} is the angle between $\overrightarrow{\mu_A}$ and $\vec{R}$. In most applicative cases, both fluorophores are assumed to be freely rotating, which results in the average value of 2/3 for κ^2. Also, for the spectral overlap, J, it is formulated by the below equation (Fig. 2.2, right):

$$J = \int f_D(\lambda)\, \varepsilon_A(\lambda)\, \lambda^4\, d\lambda$$

where $f_D(\lambda)$ is the normalized donor emission spectrum, and $\varepsilon_A(\lambda)$ is the normalized absorption spectrum or wavelength-dependent extinction coefficient of the acceptor.

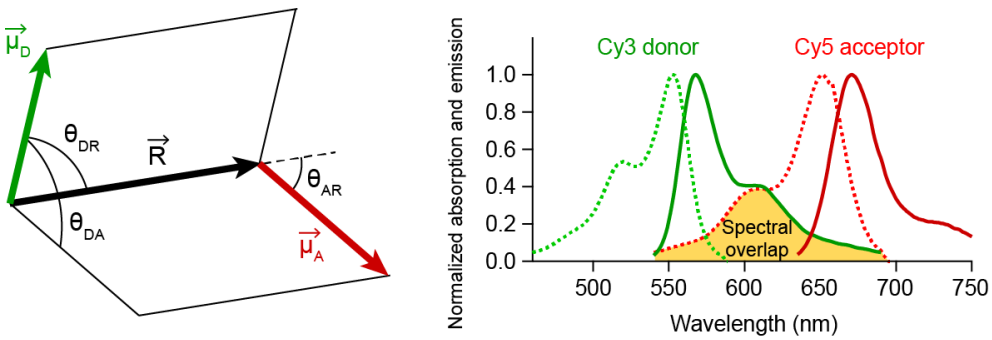

Figure 2.2. Schematics of orientation factor and spectral overlap between donor emission and acceptor absorption. The left schematics represents the relative orientations between the dipole moments of the FRET pairs and distance vector, used for orientation factor calculation. The right spectra show the definition of spectral overlap: the overlap integral of the donor emission and acceptor absorption spectra.

The FRET efficiency, E_{FRET}, is defined as the ratio of the transferred energy to total energy applied for the donor excitation, which can be expressed as:

$$E_{FRET} = \frac{\text{transferred energy to acceptor}}{\text{energy applied for donor excitation}} = \frac{k_{FRET}}{k_{FRET} + {k^D}_{fl} + \sum {k^D}_{nr}}$$

where k_{FRET}, ${k^D}_{fl}$, and ${k^D}_{nr}$ are as described before. Incorporating the equation $(*)$ for k_{FRET}, E_{FRET} can be expressed as follows:

$$E_{FRET} = \frac{1}{1 + (R/R_0)^6}$$

where R is the inter-fluorophore distance, and R_0 is the Förster radius. This equation shows that the FRET efficiency, as well as the FRET rate, depends on the

inverse sixth power of the inter-fluorophore distance and that the physical meaning of R_0 is the inter-fluorophore distance for E_{FRET} as 0.5 (Fig. 2.3).

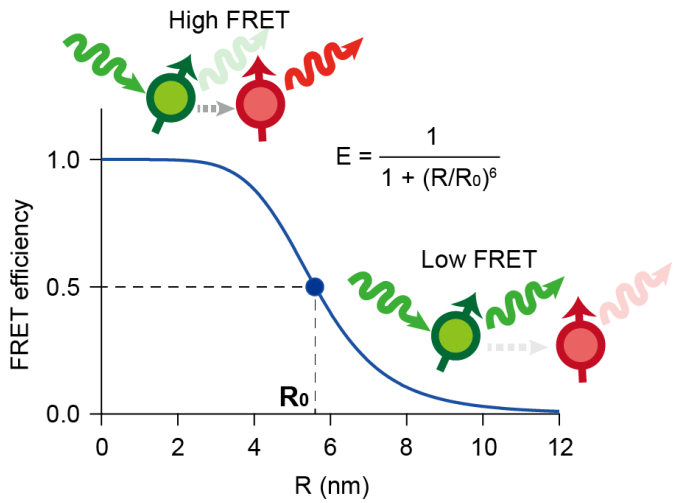

Figure 2.3. Inter-fluorophore-distance dependency of FRET efficiency. FRET efficiency depends on the inter-fluorophore distance to the inverse sixth power, and the Förster radius refers to the distance for $\boldsymbol{E_{FRET} = 0.5}$.

Experimentally, the FRET efficiency can be measured with the fluorescence lifetimes or intensities of the donor in the absence and presence of the acceptor according to the following equation, derived based on the concept that FRET serves as an additional non-radiative relaxation pathway for the excited state of the donor:

$$E_{FRET} = 1 - \frac{k^D{}_{fl} + \sum k^D{}_{nr}}{k_{FRET} + k^D{}_{fl} + \sum k^D{}_{nr}} = 1 - \frac{\tau_{D-A}}{\tau_D{}^\circ} = 1 - \frac{I_{D-A}}{I_D{}^\circ}$$

where τ_{D-A} and $\tau_D{}^\circ$ are the fluorescence lifetime of the donor in the presence and absence of the acceptor, respectively (i.e. $\tau_{D-A} = (k_{FRET} + k^D{}_{fl} + \sum k^D{}_{nr})^{-1}$ and $\tau_D{}^\circ = (k^D{}_{fl} + \sum k^D{}_{nr})^{-1}$), and I_{D-A} and $I_D{}^\circ$ are the fluorescence intensities of the donor in the presence and absence of the acceptor, respectively (i.e. $I_{D-A} = I_{ex} \cdot$

$k^D{}_{fl} \cdot \tau_{D-A}$ and $I_D{}^\circ = I_{ex} \cdot k^D{}_{fl} \cdot \tau_D{}^\circ$ where I_{ex} is the total incident light intensity for the donor excitation). The above equation can also be expressed in terms of the donor (I_D) and acceptor (I_A) fluorescence intensities when FRET occurs with the assumption that $I_D{}^\circ$ can be expressed as the sum of I_D and I_A:

$$E_{FRET} = 1 - \frac{I_{D-A}}{I_D{}^\circ} = \frac{I_A}{I_D + I_A}$$

Since most fluorescence microscope or spectrometer mainly detect the fluorescence intensity as the signals, this expression is the most widely used formula for the FRET efficiency calculation. In further experimental detail, some corrections should be applied due to the deviation of the detected raw donor intensity and the raw acceptor intensity from the assumption. In general, two correction factors, leakage factor (β) and gamma factor (γ), are introduced to calculate the true FRET efficiency:

$$E_{FRET} = \frac{(I_A - \beta I_D)}{\gamma I_D + (I_A - \beta I_D)} \quad \cdots (**)$$

where β is for the spectral leakage of donor fluorescence into the acceptor detection channel, and γ is for the calibration of the difference between donor and acceptor in fluorescence quantum yield (ϕ) and detection efficiency (η):

$$\gamma = (\eta_A/\eta_D) \times (\phi_A/\phi_D).$$

This relation, equation ($**$), is the one that we employed in data analysis for our study described in the following chapters. The values of the correction factors depend on FRET pairs as well as optical components and detectors; thus, they should be determined empirically. For Cy3-Cy5 FRET pair in our experimental setup, β was determined to be 0.15, and γ to be 1.

2.1.3. Total internal reflection fluorescence (TIRF) microscopy

Single-molecule fluorescence detection, including smFRET, requires high spatial resolution in order to discriminate a signal from a single molecule from background signals from the nearby other molecules. This can be effectively achieved by the sufficient reduction of the excitation volume. Total internal reflection fluorescence (TIRF) microscopy enables such reduction by taking advantage of an evanescent field generated at the interface where total internal reflection occurs. Thus, the TIRF microscopy has widely been applied to observe surface-tethered bio-macromolecules with fluorescent dyes labeled in real-time at the single-molecule level, thereby dissecting microscopic molecular dynamics and kinetics in the relevant biological processes.

2.1.3.1. Evanescent wave generation at the interface of total internal reflection

Total internal reflection (TIR) is an intrinsic physical property of light, occurring at the interface between the two media. According to the Snell's law, an incident light from the first medium of the refractive index of n_1 is related to the refracted light to the second medium of the refractive index of n_2:

$$n_1 \sin\theta_1 = n_2 \sin\theta_2$$

where θ_1 and θ_2 represent the incident and refracted angle, respectively (Fig. 2.4). In case of n_1 greater than n_2, the critical angle (θ_c) for the incident light, which makes the refracted angle be 90°, can be calculated:

$$\theta_c = \arcsin(\frac{n_2}{n_1})$$

This relation shows that when θ_1 is equal to θ_c, the refracted light propagates along the interface between the two media without penetrating to the second medium.

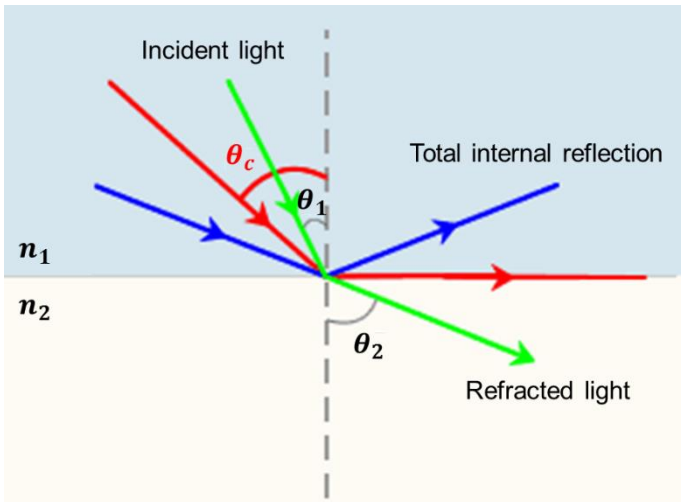

Figure 2.4. Schematics of the Snell's law and TIR. An incident light is refracted at the interface between two different media according to the Snell' law. When the refractive index of the incident medium ($\boldsymbol{n_1}$) is larger than the refracted medium ($\boldsymbol{n_2}$), total internal reflection occurs with the incident angle larger than the critical angle.

More importantly, when θ_1 is greater than θ_c, the refracted light becomes an evanescent wave propagating along the x-axis (the interface) with exponentially decaying amplitude along the z-axis (penetrating direction from the medium 1 to medium 2). The physical principle of the generation of the evanescent wave can be described as follows. First, the wavevector ($\overrightarrow{k_2}$) of the refracted wave in xz-plane is:

$$\overrightarrow{k_2} = k_2 \sin(\theta_2)\,\hat{x} + k_2 \cos(\theta_2)\,\hat{z}$$

For $n_1 > n_2$ and $\theta_1 > \theta_c$, $\sin(\theta_2)$ should be larger than 1 regarding the following relation:

$$\sin(\theta_2) = \frac{n_1}{n_2}\sin(\theta_1) > \frac{n_1}{n_2}\sin(\theta_c) = \frac{n_2}{n_2} = 1$$

Then, $\cos(\theta_2)$ becomes a complex number since $\cos(\theta_2) = \sqrt{1 - \sin^2(\theta_2)} = i\sqrt{\sin^2(\theta_2) - 1}$. Incorporating this relation to the vector expression for the electric field of the refracted wave ($\overrightarrow{E_2}$) results in:

$$\overrightarrow{E_2} = \overrightarrow{E_0}e^{i(\overrightarrow{k_2}\cdot\vec{r} - \omega_2 t)} = \overrightarrow{E_0}\exp\left[i\left\{k_2 x\sin(\theta_2) + k_2 z i\sqrt{\sin^2(\theta_2) - 1} - \omega_2 t\right\}\right]$$

where $\overrightarrow{E_0}$ is the amplitude vector of the refracted electric field, and ω_2 is the angular frequency (i.e. $\omega_2 = k_2 c/n_2$) with $\vec{r}$ and t for the position vector and time, respectively. This equation can be arranged into the following formula:

$$\overrightarrow{E_2} = \overrightarrow{E_0}e^{-\kappa z}e^{i(k' x - \omega_2 t)}$$

where $\kappa = \frac{\omega_2}{c}\sqrt{(n_1\sin\theta_1)^2 - {n_2}^2}$ and $k' = \frac{\omega_2}{c}(n_1\sin\theta_1)$. This relation is the physical description of the evanescent wave, characterized by the exponential attenuation in the z-axis. Therefore, the intensity of the evanescent wave, which is proportional to the square of the electric field (i.e. $I_2 \propto \left|\overrightarrow{E_2}\right|^2$), also decays along the penetration axis:

$$I(z) = I(0)\exp(-\frac{z}{d})$$

where d is the characteristic penetration depth depending on wavelength (λ), the refractive indices of the two media (n_1 and n_2), and the incident angle (θ_1) according to the following relation:

$$\mathrm{d} = \frac{1}{2\kappa} = \frac{\lambda}{4\pi}[(n_1\sin\theta_1)^2 - {n_2}^2]^{-\frac{1}{2}}$$

In our TIRF microscope, the excitation beam was aligned to be totally internal-reflected at quartz-water interface, in which case the penetration depth of the generated evanescent wave is calculated to be an approximately 100 nm scale. This sufficiently reduced illumination volume enabled specific excitation of the surface-immobilized molecules. Together with sparse distribution of single molecules at the surface in order to resolve each signals from the single molecules in xy-plane (i.e. the surface), the single-molecule observation using TIRF microscope was successfully achieved.

2.1.3.2. Experimental setup for the TIRF microscopy

In our studies, we employed home-built prism-type TIRF microscopy, in which a (Pellin-Broca) prism, set on the quartz slide surface, mediates TIR of the incident excitation laser beam at the quartz-water interface (Figure 2.5). The critical angle of the quartz-water interface for TIF is calculated to be 66°:

$$\theta_c = \arcsin\left(\frac{n_{water}}{n_{quartz}}\right) = \arcsin(\frac{1.33}{1.46}) \approx 66°$$

Thus, when the incident angle was aligned to be larger than 66°, the laser beam totally reflected with generating the evanescent field. To minimize unwanted refraction at the prism-quartz interface, the refractive indices of the prism and immersion oil between the prism and quartz were closely matched to that of quartz, i.e. 1.45.

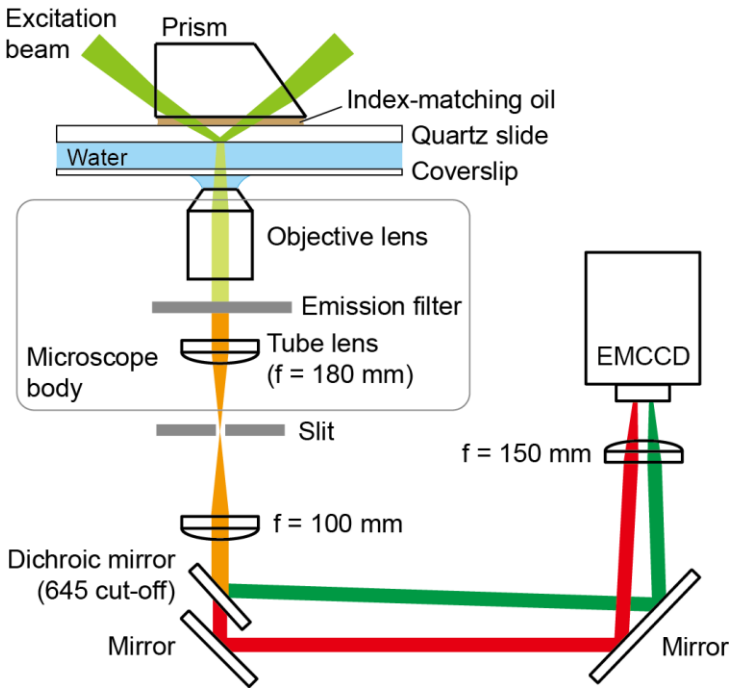

Figure 2.5. Schematics of prism-type TIRF Microscope setup. With setting a prism, excitation laser beam is aligned to be totally reflected at the quartz-water interface. Fluorescence signals from the molecules at the interface are collected by the objective lens, spectrally split by dichroic mirror, and then detected by EMCCD camera separately.

The fluorescence signals from the surface-immobilized molecules illuminated by the evanescent field were collected by a water-type objective lens (Olympus UplanSApo 60x, N.A. 1.20) and detected by EMCCD camera with green (FRET donor) and red (FRET acceptor) fluorescence split (Figure 2.5). To be specific, fluorescence signals, passing through the objective lens, were first filtered by an emission filter (ZET405/488/532/642m, Chroma) to remove Rayleigh scattering of excitation laser beams, and then focused by a tube lens in the microscope body (f = 180 mm) with magnification of 60. The signals were re-

collimated by an achromatic lens (f = 100 nm), followed by the spectral splitting via dichroic mirror (645dcxr, Chroma). The separated fluorescence signals were re-focused by another lens (f = 150 nm) on the EMCCD camera (iXon DU-897 ECS-BV, Andor) with another magnification of 1.5.

2.1.4. Enhancement of single-fluorophore photo-stability

2.1.4.1. Oxygen scavenger

Since oxygen molecules are highly reactive due to the unpaired electrons even at their electronically ground state, small organic fluorophores, commonly employed in single-molecule fluorescence studies, readily undergo photo-oxidization reaction with oxygen molecules, which permanently photo-bleaches the fluorophores. Therefore, for the long-time observation of single molecules, the oxygen species should be effectively removed in order to protect the fluorophores from the photo-bleaching. In our single-molecule study, two kinds of enzyme-utilized oxygen scavenging systems were employed: glucose oxidase and catalase (GOC)[22] and pyranose oxidase and catalase (POC)[23]. The reaction mechanism of these enzymatic systems are as follows:

$$\beta\text{-D-glucose} + O_2 + H_2O \xrightarrow{\text{glucose oxidase}} \text{gluconic acid} + H_2O_2$$

$$\beta\text{-D-glucose} + O_2 + H_2O \xrightarrow{\text{pyranose oxidase}} \text{2-dehydro-D-glucose} + H_2O_2$$

$$H_2O_2 \xrightarrow{\text{catalase}} H_2O + \frac{1}{2}O_2$$

These two systems both removes dissolved oxygen molecules effectively, but the

difference is that the GOC system drops the pH of the solution whereas the POC system does not, by which POC becomes more suitable for the long-term observation, particularly, of pH-sensitive systems, e.g. biological systems.

2.1.4.2. Triplet quencher

Although inducing photo-bleaching of the organic fluorophores, the molecular oxygen effectively quenches the triplet states of the fluorophore to the electronically ground state, thereby reducing photo-blinking, which occurs when the excited fluorophore becomes trapped at the long-lived, non-fluorescent triplet states for, typically, μs ~ ms time scale. Thus, with the removal of the molecular oxygen via the oxygen scavengers, a triplet quencher (typically a reducing agent mixed with its oxidized form) should be incorporated in order to suppress the photo-blinking.

In our single-molecule fluorescence study, we used Trolox ((±)-6-hydroxy-2,5,7,8-tetramethylchromane-2-carboxylic acid), the most widespread triplet quencher employed in single-molecule fluorescence detection. The Trolox solution, prepared by the partial photo-oxidization of Trolox to Trolox-quinone, quench the triplet states of fluorophores effectively via rapid redox reactions coupled with Trolox (a reductant) and Trolox-quinone (an oxidant)[24] (Fig. 2.6). Therefore, the triplet quencher along with the oxygen scavenging system dramatically enhances the photo-stability of small organic fluorophores in terms of both photo-bleaching and blinking.

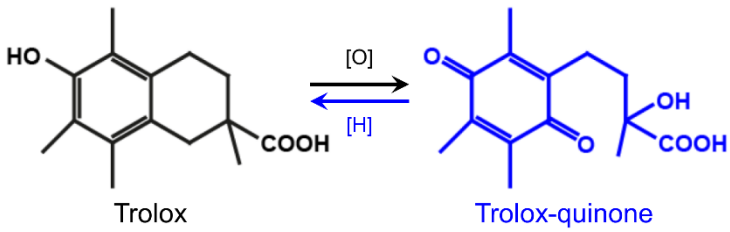

Figure 2.6. Chemical structural formulas for Trolox and Trolox-quinone. Trolox and Trolox-quinone undergo numerous cycles of interconverting redox reactions, thereby quenching triplet states of excited organic fluorophores.

2.2. CRISPR/Cas9 System

CRISPR (clustered regularly interspaced short palindromic repeats) / Cas9 (CRISPR-associated protein 9) system is one type of prokaryotic adaptive immune system, which distinguishes and degrades foreign DNA. The system employs only one protein component, Cas9, to structure a functional ribonucleoprotein effector complex with two guide RNAs (gRNAs), CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA)[7], in which tracrRNA holds Cas9 and crRNA together by base-pairing and RNA-protein interaction, respectively (Fig. 2.7). The Cas9-gRNA effector complex (Cas9:gRNA) binds to the target DNA sequences via recognition of short nucleotide sequences called protospacer-adjacent motif (PAM) in the target DNA and the complementary base-pairing between crRNA and the target DNA[8–10]. In this section, we summarize the previous reports on the mechanism of action of the CRISPR/Cas9 system, focusing on the gRNAs.

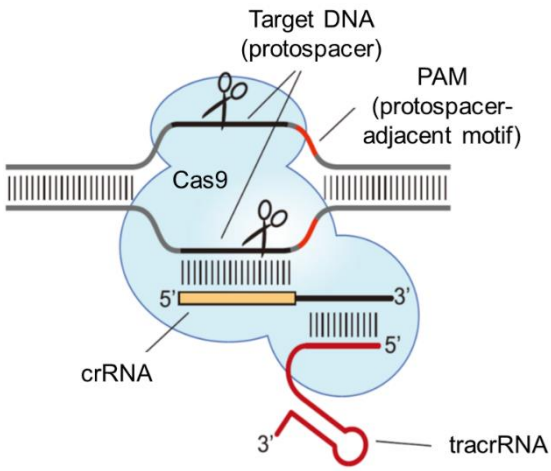

Figure 2.7. Cartoon for Cas9:gRNA:DNA complex. CRISPR/Cas9 system functions in the form of Cas9:gRNA complex, which consists of Cas9 endonuclease with two guide RNAs (crRNA and tracrRNA). Cas9:gRNA binds to target DNA via PAM interaction and crRNA-DNA base-pairing and then cleaves the DNA with nuclease domains. The figure is modified from the ref. 10.

2.2.1. Structural rearrangement of Cas9 induced by guide RNA

Most structural characteristics of *apo*-Cas9, Cas9:gRNA, and Cas9:gRNA:DNA complex have been revealed by X-ray crystallography and electron microscopy, one of which is the global structural rearrangements of Cas9 protein domains upon gRNA and DNA binding (Fig. 2.8). Specifically, the single-particle electron microscopy showed that the nuclease lobe, composed of nuclease domains and PAM-interacting (PI) domain, rotates upon gRNA binding in order to open the DNA-binding channel, buried deep inside the *apo*-Cas9 structure, and re-closes upon DNA binding to hold and cleave the DNA[19]. The X-ray crystallographic study further revealed the details on the conformational rearrangement that the PI domain pre-organizes with gRNA

binding primed for PAM recognition accompanied by the PAM-proximal pre-organization of gRNA primed for DNA-RNA heteroduplexation[18]. With crystal structures of Cas9:gRNA:DNA[20], the HNH nuclease domain was revealed to move for cleavage competency from the PAM-distal region toward PAM-proximal cleavage site upon gRNA and DNA binding. These structural reports indicate that the Cas9 protein needs multi-steps for structural activation, which is closely related to gRNA.

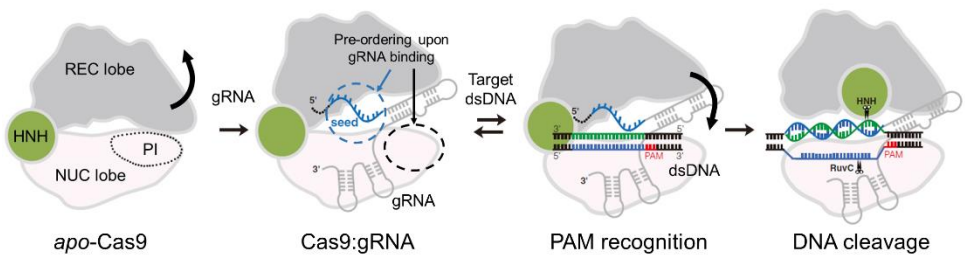

Figure 2.8. Structural rearrangement of Cas9 domains coupled with incorporation of gRNA and DNA. Cas9 undergoes multiple structural rearrangements during the Cas9:gRNA:DNA complex formation. The figure is modified from the ref. 18.

2.2.2. Mechanism of gRNA-DNA heteroduplexation in Cas9:gRNA:DNA complexation

One of the most distinctive features of the Cas9 endonuclease is that it lacks a helicase domain; it unwinds target double-stranded DNA (target dsDNA) without consumption of ATP. The thermal-energy-derived unwinding of the target dsDNA, thus, largely relies on the heteroduplexation of complementary strand of the target dsDNA with crRNA, which stabilizes the unwound single-stranded DNA. Biochemical competition assays suggested that the dsDNA melting and crRNA-

DNA heteroduplexation, i.e. R-loop structure formation, proceed unidirectionally after the PAM recognition, from the PAM-proximal bases to the PAM-distal ones[15] (Fig. 2.9). Other studies employing ChIP (Chromatin immunoprecipitation)[25] and magnetic tweezer[26] corroborated the hypothesis of unidirectional R-loop expansion, and further revealed the importance of the 'seed' nucleotides (approximately 6 ~ 12 bases from the PAM-proximal end) for the formation of the stable Cas9:gRNA:DNA ternary complex; off-target DNAs containing fully matched seed bases with mismatched PAM-distal bases recruited Cas9:gRNA effectively. These discoveries imply the structural flexibility on the PAM-distal region in the R-loop structure.

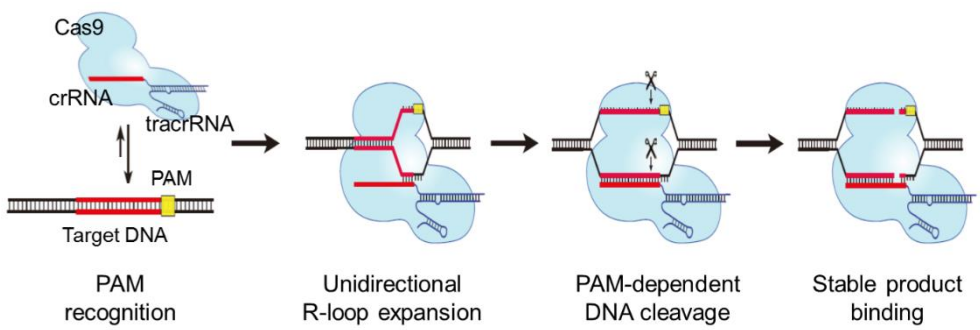

Figure 2.9. Unidirectional heteroduplexation as the mechanism of Cas9:gRNA-dsDNA binding. Cas9:gRNA complex binds to target dsDNA with the directionality from the PAM-proximal end to PAM-distal end. The figure is modified from the ref. 15.

2.2.3. Nuclease activity of Cas9 related to PAM-distal complementarity between crRNA and DNA

In contrast to the above reports, *in vivo* study on nuclease activity of the Cas9 complex reported that crRNA with 1 ~ 3-nt truncation from the PAM- distal end retains catalytic activity of Cas9:gRNA whereas more than 3-nt removal abolishes

the activity nearly to zero[27]. In particular, the nuclease activity dropped dramatically within only 1-nt difference (i.e. ~90% activity for 3-nt truncation versus ~0% activity for 4-nt truncation), indicating that the activity collapse is not attributed to gradual destabilization of the crRNA-DNA heteroduplex due to the shortening of the base-pairing length. Furthermore, recent ensemble-FRET spectroscopic study revealed that the PAM-distal complementarity of DNA to crRNA directly affects the nuclease activity of the Cas9 complex along with structural activation of the HNH nuclease domain[28] (Fig. 2.10). Therefore, these previous reports suggest that the PAM-distal heteroduplexation between crRNA and DNA serves as an allosteric regulator for the structural activation of the HNH nuclease domain of Cas9. Given the afore-discussed structural flexibility of crRNA on the PAM-distal region, sub-populational analysis and molecular dynamics study are required to understand the molecular-level association between the PAM-distal flexibility and the allosterically regulated nuclease activity of the Cas9 complex.

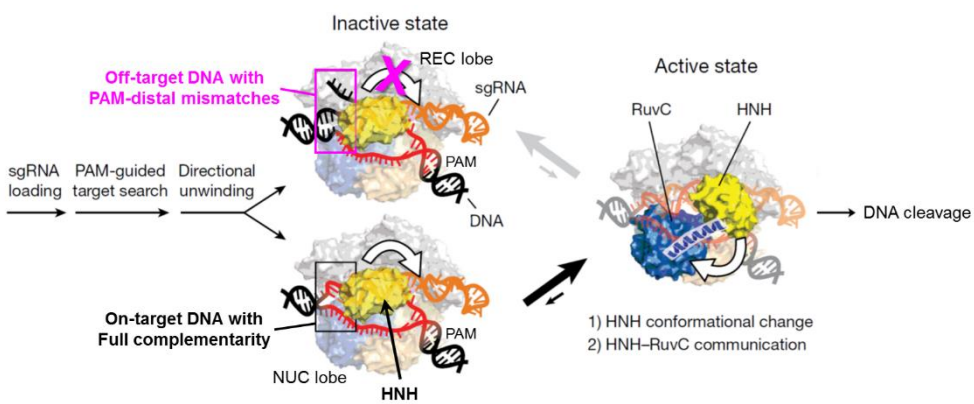

Figure 2.10. Structural activation of HNH-nuclease domain by PAM-distal complementarity. PAM-distal mismatches between crRNA and DNA abrogates HNH-nuclease-domain movement and thereby the nuclease activity. The figure is modified from the ref. 28.

3. Inhibitory Effect of tracrRNA on Conformational Inactivation of Cas9

3.1. Introduction

For the formation of CRISPR/Cas effector complex, type-II CRISPR/Cas9 system requires tracrRNA as well as crRNA, whereas the other types only need crRNA as gRNA. According to X-ray crystallographic studies[17,18], tracrRNA interacts extensively with Cas9 throughout its full length, proving its importance in stabilization of functional Cas9:gRNA. Furthermore, *in vivo* biochemical approach reveals that tracrRNA binds to pre-crRNA, transcribed from CRISPR locus in genomic DNA, thereby inducing crRNA-maturation mediated by Cas9 and RNaseIII[5]. Although these previous reports showed that tracrRNA is critical for the formation of mature Cas9:gRNA, the sequential process of crRNA-tracrRNA-Cas9 complexation was still unknown; whether (pre-)crRNA hybridizes to tracrRNA first, followed by Cas9 binding or (pre-)crRNA binds to already-formed Cas9-tracrRNA binary complex. Thus, this chapter deals with our investigation into the sequential assembly of Cas9, tracrRNA, and crRNA for the formation of functional Cas9:gRNA, monitored by single-molecule DNA-cleavage assay, and the newly unveiled structural role of tracrRNA in the formation process of Cas9:gRNA.

3.2. Materials and Methods

Materials – DNA, Cas9, and RNA: DNA oligonucleotides for the target DNA were purchased from Integrated DNA Technologies. 5'-end biotinylation was introduced to non-complementary strand for surface-immobilization, and internal amino modification for fluorescent dye labeling was incorporated in both DNA

strands. Cy3 and Cy5 fluorophores were conjugated to the modified amino group of the non-complementary strand and complementary strand, respectively, by nucleophilic substitution reaction between amino-modified thymine and NHS ester-linked fluorescent dyes according to the well-established protocol[29]. Double-stranded target DNA was prepared by annealing the Cy3-linked (biotinylated) non-complementary strand with (Cy5-linked) complementary strand at the molar ratio of 1:1.2 (50μM for the non-complementary strand) in buffer condition of 25 mM Tris pH 8.0 and 250 mM NaCl. The annealing was conducted by heating the mixture to 95 °C for 2 min, followed by slow cool-down for approximately 3 hr.

Cas9 protein from S. pyogenes (SpyCas9) was prepared by overexpression from E. coli. Recombinant SpyCas9 gene containing a nuclear localization signal, HA epitope, and His-tag at the N-terminus was subcloned into pET28-b(+), then overexpressed in BL21(DE3) strain, purified using Ni-NTA agarose beads (Qiagen), dialyzed against 20 mM HEPES pH 7.5, 150 mM KCl, 1 mM DTT, and 10% glycerol, and finally, concentrated using Ultracel 100K cellulose column (Millipore). The purity and concentration of Cas9 protein were quantified by SDS-PAGE.

Guide RNAs (both tracrRNA and crRNA) were prepared by *in vitro* transcription using MEGAshortscript T7 kit (Ambion) according to the manufacturer's manual. Transcribed RNAs were purified by phenol:chloroform extraction, chloroform extraction, followed by ethanol precipitation. The concentration of the purified RNAs was quantified by UV-Vis spectrometry. All RNA-containing solutions were prepared with nuclease-free water with all disposable tips and microcentrifuge tubes qualified as RNase-free and all

experimental devices and lab space treated with 70% ethanol.

All the RNA and DNA sequences with the positions of biotin and internal amino modifications are listed in Appendix.

Single-molecule DNA-cleavage assay: For each single-molecule observation, a flow chamber composed of a microscope slide and a coverslip, sealed with double-sided tapes and epoxy, was assembled. Two tiny holes were made on the microscope slide before the assembly of the flow chamber so that they work as an inlet and outlet for solution exchange. All coverslips and quartz slide glasses were passivated by poly-ethylene-glycol (PEG) to prevent nonspecific binding of samples to the glass surface[30]. All imaging and DNA-cleavage reaction were performed at 37 °C (unless otherwise specified) in the buffer composition of 100 mM NaCl, 50 mM Tris-HCl pH 7.9, 10 mM MgCl2, 1 mM DTT and 0.1 mg/ml BSA with the oxygen scavenger (2.7 U/ml of pyranose oxidase (Sigma-Aldrich), 7.5 U/ml of catalase (Sigma-Aldrich) and 0.4% (w/v) of D-glucose) and the triplet quencher (~2 mM Trolox) in order to enhance photostability of the organic fluorophores[23].

For the single-molecule cleavage assay, the DNA-cleavage efficiencies were measured after 5-min incubation of 2 nM Cas9 and 300 nM gRNAs with the surface-immobilized DNA unless otherwise stated.

Set-up and data analysis: We conducted all single-molecule cleavage assays with home-built total internal reflection fluorescence microscopy described in Chapter 2. A 532 nm green laser (Samba, Cobolt AB) and 633 nm He-Ne laser (25-LHP-925-

230, Melles-Griot) were employed simultaneously to excite Cy3 and Cy5 molecules, respectively. The detected fluorescence images were recorded by a homemade C++ script (available for download at https://cplc.illinois.edu/software).

The ratio of cleaved DNA was calculated by comparing the quantified number of Cy3 and Cy5 molecules on the DNA before the injection of the reconstituted Cas9:gRNA, and after 7 M urea-treatment to the flow chamber. IDL and MATLAB codes were built to count the number of Gaussian-fitted spots from the recorded images and then to calculate the average of (the number of Cy3 molecules)/(the number of Cy5 molecules) in order to take photo-bleaching and/or urea-mediated denaturation of the DNA duplex into account. To obtain DNA-cleavage efficiency at pseudo-ensemble level, data from more than 10 measurements were collected with each measurement recording signals from over 300 molecules simultaneously.

Circular dichroism (CD) measurement: CD spectra of *apo*-Cas9 protein were obtained using Chirascan plus (Applied Photophysics) over the range of 190 to 260 nm at indicated temperatures. The data acquisition was performed with 1 nm-interval between 190 and 260 nm with triplicate wave-scans, and Peltier-type temperature controller controlled the sample temperature from 25 to 37 ºC. Concentration of the *apo*-Cas9 sample for the measurement was 1 mg/ml in buffer composition of 20 mM HEPES pH 7.5, 150 mM KCl, 1 mM DTT, and 10% glycerol.

Thermal melting curve measurement: Thermal melting curve of *apo*-Cas9 and Cas9-crRNA mixture were obtained by monitoring the change in the ratio of intrinsic tryptophan fluorescence intensity at the emission wavelengths of 330 nm and 350 nm with fixed excitation wavelength at 280 nm using fluorescence spectrometer (Quantamaster, Photon Technology International), while raising temperature. Concentration of the *apo*-Cas9 for the tryptophan fluorescence measurement was 1 μM, and the concentration of crRNA was 1.5 μM in buffer composition of 10 mM Tris-HCl pH 7.4, 300 mM NaCl, 1 mM DTT, 10% glycerol.

3.3. Results and Discussion

3.3.1. Visualization of Cas9 nuclease activity at single-molecule level

To observe nuclease activity of Cas9:gRNA at single-molecule level, total internal reflection fluorescence microscopy was employed. We designed biotinylated 48-bp double-stranded DNA (dsDNA) dual-labeled with Cy3 and Cy5 fluorophores at the vicinity of 5'-end of each strand respectively and tethered to passivated surface via biotin-neutravidin interaction, enabling single-molecule detection of dual-labeled dsDNA (Fig. 3.1a for the scheme and Appendix for the sequence).

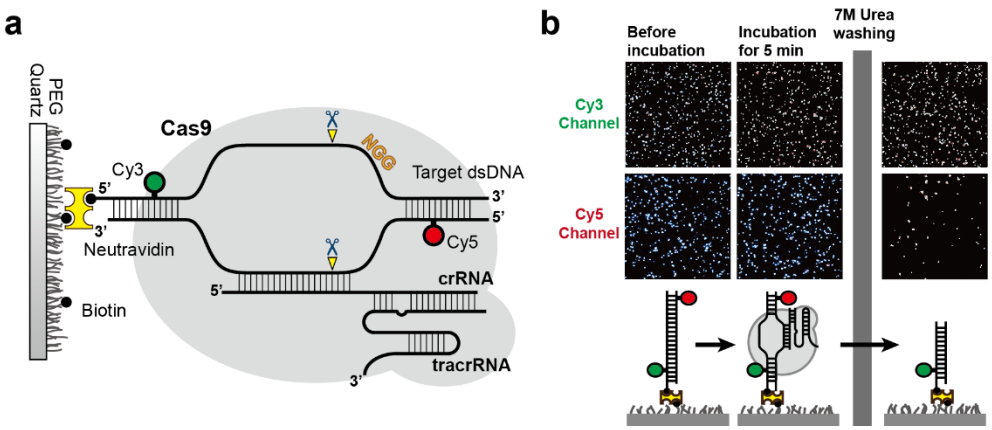

Figure 3.1. Single-molecule fluorescence visualization of DNA cleavage by Cas9:gRNA. (a) Schematics of surface-immobilized dual-labeled DNA (the distance between Cy3 and Cy5 is far enough for negligible FRET), targeted and cleaved by Cas9:gRNA. (b) CCD images showing single-molecules of the DNA before and after the cleavage reaction of Cas9:gRNA.

Cas9 (2 nM) was pre-incubated with excess gRNAs (300 nM) for 20 min at 37 °C to form Cas9:gRNA, followed by the injection of the Cas9-gRNA mixture onto the surface-immobilized target DNA. The DNA-cleavage reaction by Cas9:gRNA, through dissociation of cleaved DNA fragments, was monitored after the injection of Cas9:gRNA by calculating the ratio of the number of Cy5 molecules to that of Cy3 molecules. Consistent with a previous study reporting that Cas9:gRNA remains bound to the DNA even after the DNA cleavage[15], no change in the number of Cy3 and Cy5 molecules was detected upon the incubation of Cas9:gRNA to the DNA at 37 °C (Fig. 3.1b). To explicitly observe the dissociation of the cleaved DNA fragment and Cas9:gRNA from the immobilized fraction of the DNA, 7 M-urea solution was injected, as previously applied[15], followed by rapid wash-out in order to prevent urea-mediated denaturation of biotin-neutravidin interaction or DNA

double-helix structures which could mislead the calculated cleaved-DNA ratio. After the urea-treatment, the ratio of the number of Cy5 to Cy3 molecules decreased significantly, indicating that the immobilized target DNA was efficiently cleaved by Cas9:gRNA; approximately 80% of target DNAs were cleaved in 5 min (Fig. 3.1b and Fig. 3.3). However, when Cas9 was replaced by dCas9 (a catalytically 'dead' Cas9 by introducing point-mutations at the catalytically active sites, D10A and H840A) as a negative control, negligible fraction of DNAs were cleaved (Fig. 3.2), verifying that the observed 80% loss in Cy5 molecules are predominantly attributed to the DNA-cleavage activity of Cas9:gRNA.

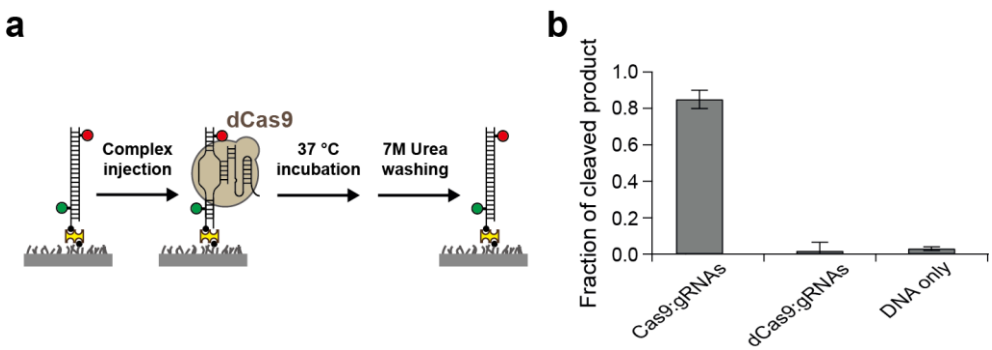

Figure 3.2. Single-molecule DNA-cleavage assay using dCas9. As a control, the single-molecule DNA-cleavage assay with dCas9 instead of catalytically active Cas9 results in the cleaved DNA ratio at background level (i.e. DNA only case) (mean ± SEM, n ≥ 5).

By varying incubation time of Cas9:gRNA with the target DNA, Cas9-induced DNA-cleavage kinetics was investigated. The enzymatic kinetic curve was well-fitted by a single-exponential fit (Fig. 3.3), suggesting that the reaction mechanism of the target-DNA cleavage by Cas9:gRNA contains a first-order rate-determining step (RDS). The result that the decay constant of the single-exponential

fit varies with the concentration of Cas9 implies that the pseudo-first-order rate constant depends on the concentration of Cas9, indicating that the binding of Cas9:gRNA to the target DNA is the RDS for Cas9:gRNA functioning.

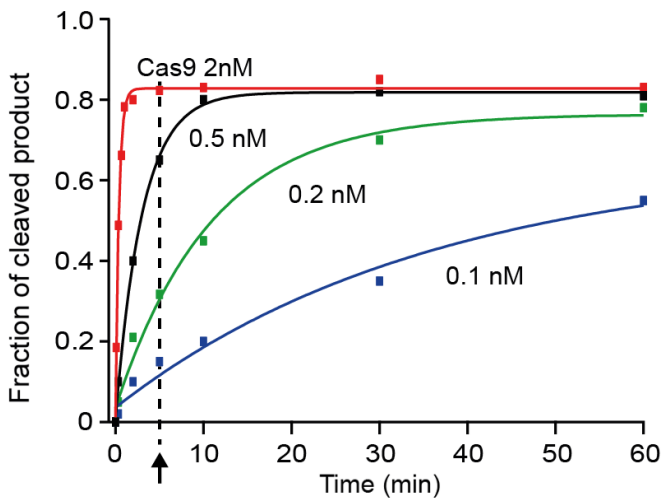

Figure 3.3. Kinetic curves of Cas9-mediated DNA-cleavage reaction at various Cas9 concentrations. Kinetic analysis of the cleavage reaction shows single-exponential kinetics, in which the rate constant depends on concentration of Cas9. With 2 nM of Cas9, around 80% of DNA molecules were cleaved in 5 min (black arrow).

3.3.2. Identification of inactivated *apo*-Cas9 in the absence of tracrRNA

To explore the sequential roles of gRNAs in the DNA-cleavage activity, we compared the DNA-cleavage efficiency from different sets of partially pre-incubated Cas9-gRNAs mixtures using the single-molecule DNA-cleavage assay described above. Three different sequential processes of pre-incubation and mixing scheme were employed before the injection of the mixture onto the immobilized DNA (Fig. 3.4a): (i) *apo*-Cas9 and tracrRNA pre-incubated at 37 °C for 20 min, and crRNA

added afterwards; (ii) *apo*-Cas9 and crRNA pre-incubated at 37 °C for 20 min, and tracrRNA added afterwards; and (iii) *apo*-Cas9 alone pre-incubated at 37 °C for 20 min, and both tracrRNA and crRNA added afterwards. Additionally, *apo*-Cas9 pre-incubated at 37 °C for 20 min with both gRNAs ('full complex') was employed as a positive control. The cleavage efficiency was then measured at 37 °C after an DNA-Cas9:gRNA incubation time of 5 min, the time enough to reach maximum cleavage under our experimental conditions (Fig. 3.3).

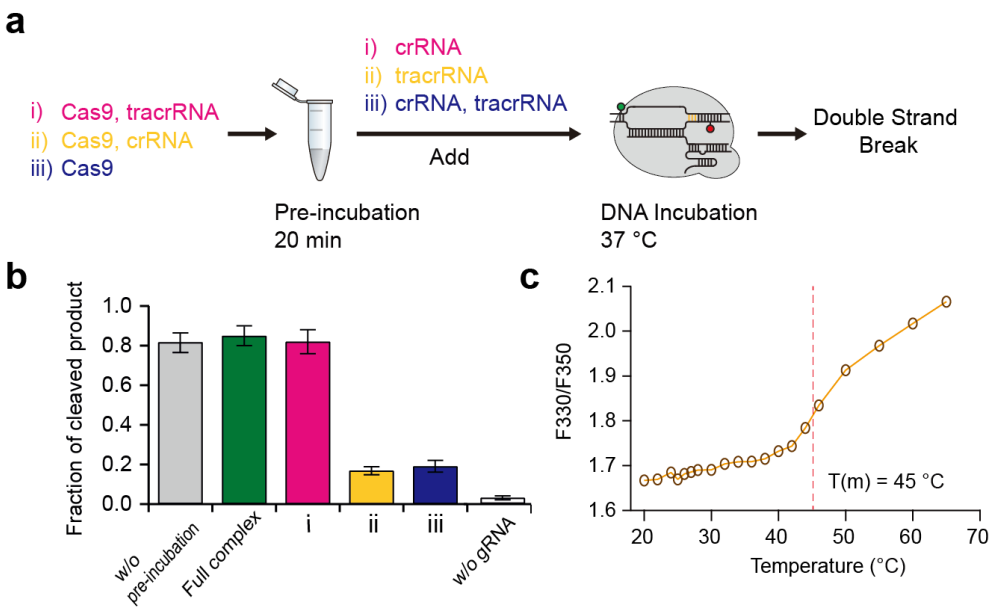

Figure 3.4. Inactivation of *apo*-Cas9 in the absence of tracrRNA. Schematics of the experimental processes (a) and DNA-cleavage results (mean ± SEM, n ≥ 5) (b) for partial pre-incubation experiments. (c) Thermal melting curve of Cas9-crRNA mixture shows phase transition at 45 °C, identical to the melting temperature of *apo*-Cas9 (Fig. 3.5c).

In the case of Cas9 partially pre-incubated with tracrRNA ((i) in Fig. 3.4), the efficiency of DNA cleavage was identical to that of the 'full complex', both of which reached approximately 80% (Fig. 3.4b, magenta and green bars). Also, as a control, the cleavage efficiency for the direct injection of the mixture with *apo*-Cas9, crRNA, and tracrRNA to the DNA ('without pre-incubation', Fig. 3.4b, gray bar) was almost the same with the cleavage efficiency of the 'full complex', indicating that the 5-min incubation time is actually far enough for both Cas9:gRNA complexation and DNA-cleavage reaction to take place. In sharp contrast, however, Cas9 pre-incubated in the absence of tracrRNA ((ii) and (iii) in Fig. 3.4) yielded much lower cleavage efficiencies around 20% (Fig. 3.4b, yellow and blue bars). Considering that Cas9:gRNA:DNA remains stable even after the DNA cleavage, i.e. Cas9:gRNA does not undergo multiple turnover cycles in our experimental time scale, the different DNA-cleavage efficiencies correspond to the different concentrations of functionally active Cas9:gRNA. Thus, these results suggest that *apo*-Cas9 becomes inactivated when pre-incubated in the absence of Cas9-tracrRNA interaction, certifying tracrRNA as a prerequisite for the formation of functionally active Cas9:gRNA. (In the case of Cas9 pre-incubated with crRNA, one could raise the possibility that Cas9 is inactivated by the crRNA-Cas9 interaction, but no stable crRNA-Cas9 interaction was observed when comparing the thermal melting curve measured with the mixture of Cas9 and crRNA versus *apo*-Cas9 alone (Fig. 3.4c and Fig. 3.5c).)

DNA-cleavage efficiency of Cas9 pre-incubated without tracrRNA (process (ii) in Fig. 3.4a) also strongly depended on the pre-incubation temperature; the cleavage efficiency decreased as the pre-incubation temperature increased (Fig. 3.5a). The result suggests that the transition of functionally active *apo*-Cas9 into inactivated *apo*-Cas9 requires to overcome an energy-barrier pathway. Temperature-dependent circular dichroism (CD) spectra show that the energy-consuming transformation from active *apo*-Cas9 to inactivated *apo*-Cas9 is accompanied by conformational change (Fig. 3.5b). As the pre-incubation temperature increased from 25 to 37 °C, the ellipticity of *apo*-Cas9 in the characteristic band (210 ~ 220 nm) changed accordingly (from the black to the green curve). By contrast, when *apo*-Cas9 is annealed slowly from 37 to 25 °C, no reverse change was detected (the dashed red curve), indicating that the reverse conformational rearrangement from the inactivated state to the active one is not allowed thermodynamically. To examine whether the observed conformational transition is derived from unfolding or phase transition of *apo*-Cas9 or not, we obtained thermal melting curve of *apo*-Cas9 by measuring the shift of intrinsic tryptophan fluorescence (Fig. 3.5c). The melting temperature of *apo*-Cas9 was calculated to be 48.5 °C with only 2% of *apo*-Cas9 denatured at 37 °C, verifying that the conformational transition from active *apo*-Cas9 to inactivated *apo*-Cas9 is not related to thermal denaturation or phase transition of *apo*-Cas9. Taken together, these results suggest that there exists another conformational state of *apo*-Cas9, inactivated *apo*-Cas9, which is thermodynamically more stable than the active one, but the structural transition toward which is kinetically blocked in the absence of sufficient thermal energy.

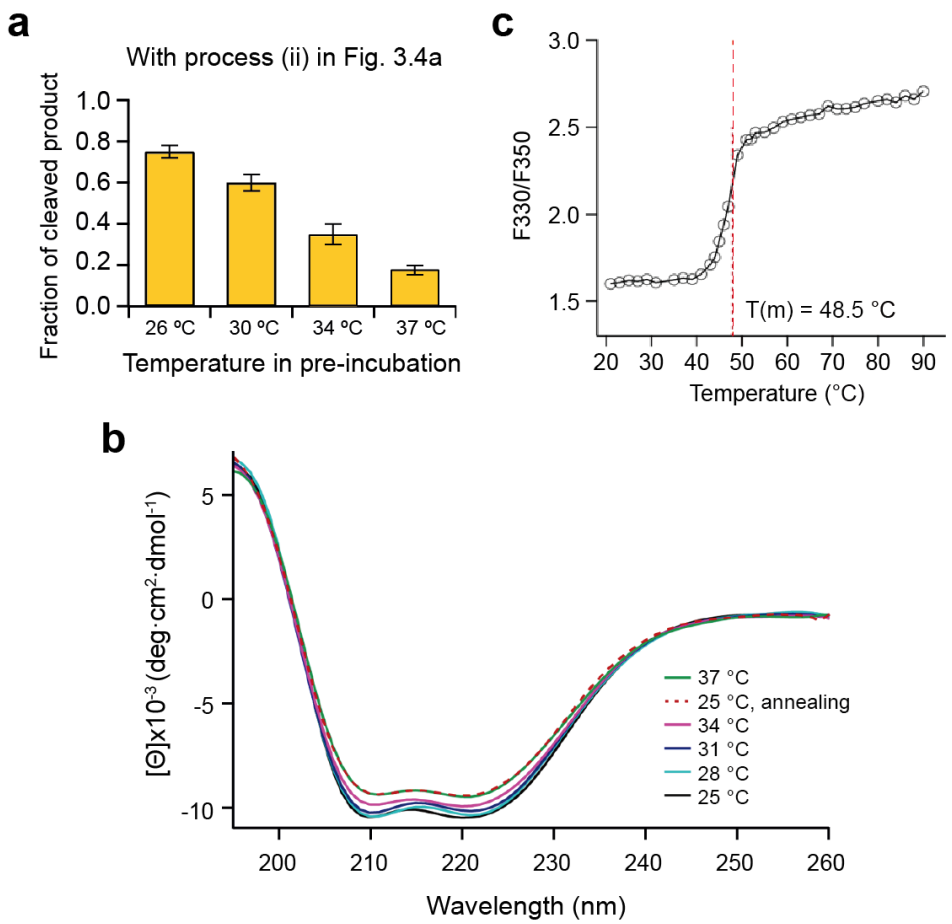

Figure 3.5. Temperature-dependency of the *apo*-Cas9 inactivation. (a) Quantification of DNA-cleavage ratio with increasing pre-incubation temperature for process (ii) in Fig. 3.4a (mean ± SEM, n ≥ 5). (b) CD spectra of *apo*-Cas9 at various temperatures. (c) Thermal melting curve of *apo*-Cas9.

3.3.3. tracrRNA-involved recovery to functionally active Cas9:gRNA

Next, to determine whether inactivated *apo*-Cas9 binds to gRNAs to form functionally active Cas9:gRNA, we ran the cleavage assay with annealed *apo*-Cas9 prepared as in Fig. 3.5b (dotted curve) after another 1-hr pre-incubation of annealed *apo*-Cas9 with two gRNAs at 22 °C (Fig. 3.6a). The result was the DNA-cleavage efficiency of only around 20%, similar to those from (ii) and (iii) in Fig. 3.4b (Fig. 3.6b). Considering the aforementioned result that inactivated *apo*-Cas9 is trapped at lower temperatures, this data shows that tracrRNA does not interact with inactivated *apo*-Cas9 to catalyze reverse conformational rearrangement of inactivated *apo*-Cas9 toward active *apo*-Cas9 or toward the formation of functionally active Cas9:gRNA at least without sufficient thermal energy.

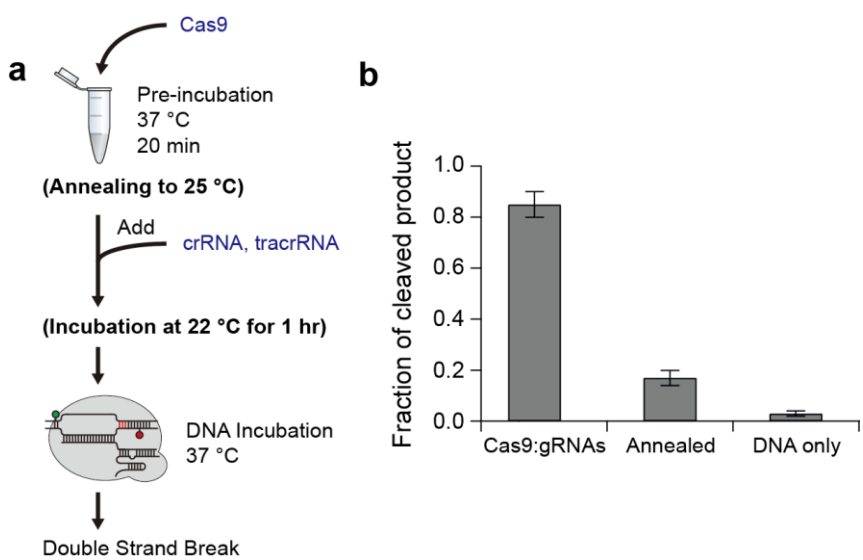

Figure 3.6. DNA-cleavage assay with annealed *apo*-Cas9. (a) Schematics of experimental process for the DNA-cleavage assay with annealed *apo*-Cas9. (b) The DNA-cleavage efficiency of annealed *apo*-Cas9 was comparable to that of pre-incubated *apo*-Cas9 at 37 °C despite of 1-hr pre-incubation time with gRNAs at 22 °C (mean ± SEM, $n \geq 5$).

We revisited the cleavage assay with inactivated Cas9, using longer DNA incubation time for the thorough investigation of the rearrangement kinetics from inactivated *apo*-Cas9 to functionally active Cas9:gRNA at the elevated temperature (37 °C). The pre-formed inactivated *apo*-Cas9 species as in Fig. 3.4a (ii) and (iii) were incubated with the immobilized DNA right after the addition of the other gRNA components, i.e. tracrRNA for (ii) and both tracrRNA and crRNA for (iii). The result shows that inactivated *apo*-Cas9 recovered its full cleavage activity after about 20 min of DNA-incubation in both cases (Fig. 3.7), implying that the thermal energy at the elevated temperature is sufficient to overcome the energy barrier for the conformational rearrangement which blocks the rearrangement at the lower temperature (22 °C). Kinetically, the cleavage efficiency started to arise after a lag phase of approximately 10 min under our experimental condition. The presence of the lag phase suggests the existence of slow-rate steps in the reaction mechanism, which could be attributed to the large-energy-barrier conformational rearrangement of inactivated *apo*-Cas9 involving tracrRNA. Therefore, these results reveal that the Cas9-tracrRNA interaction prevents Cas9 from transitioning into inactivated *apo*-Cas9 by shifting the chemical equilibrium away from the inactivated conformation by allowing a lower energy state, functionally active Cas9:gRNA (Fig. 3.8).

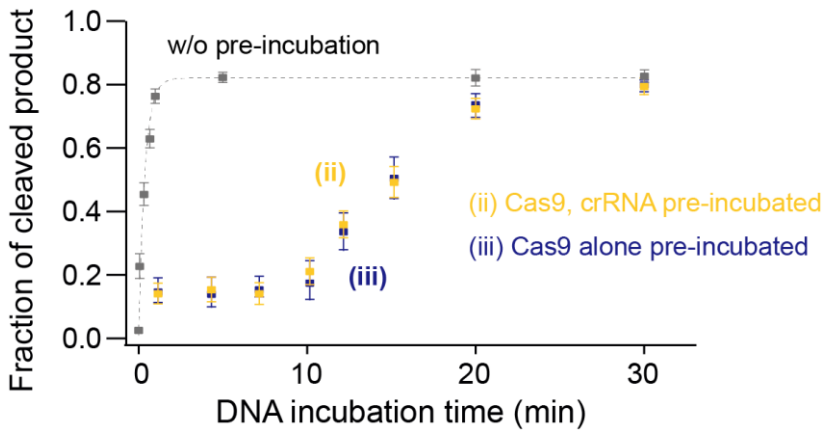

Figure 3.7. Kinetic curves for the recovery of DNA-cleavage activity of inactivated *apo*-Cas9. The DNA-cleavage efficiencies of inactivated *apo*-Cas9 species (with the pre-incubation processes (ii) and (iii) in Fig. 3.4a) were recovered after extended DNA-incubation time in the presence of tracrRNA with sufficient thermal energy (error bars represent s.d.).

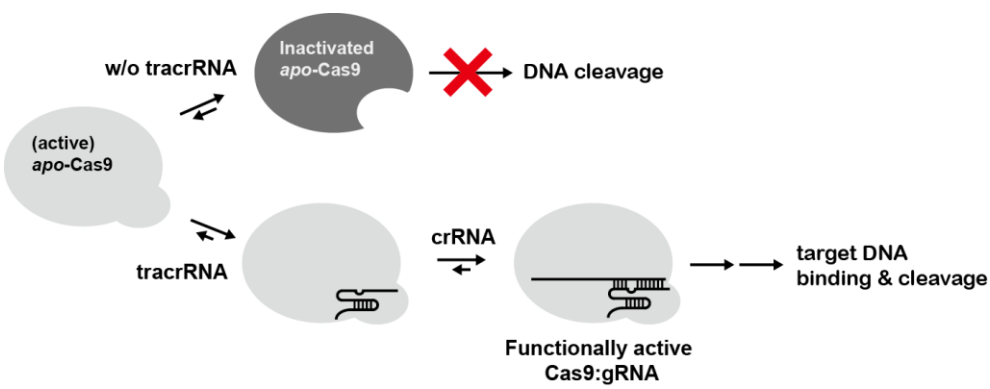

Figure 3.8. Schematic model for the tracrRNA-mediated formation pathway of functionally active Cas9:gRNA.

3.4. Conclusion

In this study, the role of tracrRNA in the formation of the functionally active Cas9 effector complex was investigated using the single-molecule DNA-cleavage assay. We demonstrated that tracrRNA is required to retain the suitable conformation of *apo*-Cas9 for functional Cas9-gRNA complexation at the physiological temperature (37 °C). Unlike previously reported conformational rearrangement upon gRNA binding[18] and target DNA hybridization[16,17,20], we found another type of conformational transition in *apo*-Cas9 itself without gRNA or DNA interactions although its *in vivo* relevance remains to be addressed by further studies. The fact that *apo*-Cas9 becomes inactivated in the absence of tracrRNA suggests that active *apo*-Cas9 first recognizes tracrRNA prior to crRNA binding during the formation process of Cas9:gRNA. This *in vitro* result could shed some light on the role of Cas9 and tracrRNA in the crRNA-maturation process, described in the introduction, in which tracrRNA may be recognized by Cas9 and then the Cas9-tracrRNA complex be recruited to pre-crRNA, similar to the Cas9-gRNA complexation process established in this study.

4. Regulatory Effect of Sub-conformational Transition of crRNA on Nuclease Activation of Cas9 Complex

4.1. Introduction

In CRISPR/Cas adaptive immunity, crRNA is a key component owing to its exclusive role in target specificity; it specifically binds to target DNA by base-pairing. In type-II CRISPR/Cas9 system, several studies have proposed that the base-pairing between crRNA and DNA forms unidirectionally. The unidirectional R-loop expansion from PAM-proximal end to PAM-distal end was first suggested based on biochemical competition assay[15] and supported by single-molecule observation of R-loop formation using a magnetic tweezer[26,31]. However, the intrinsic dynamics of the R-loop expansion was still enigmatic; the biochemical assay lacked dynamic information, and the employment of the magnetic tweezer distorted the intrinsic dynamics due to the exerted force. Hence, we looked deep into the innate dynamic properties of the R-loop expansion between crRNA and the target DNA. This chapter discusses our single-molecule observation of the R-loop structure, subsequent discovery of newly identified R-loop sub-conformations, and the regulatory role of the sub-conformational transition in nuclease activation of Cas9:gRNA:DNA.

4.2. Materials and Methods

Materials - Cas9, RNA, and DNA: Cas9 and dCas9 from S. pyogenes were overexpressed from E. coli or purchased from New England Biolabs. In the case of home-expressed proteins, the sample preparation procedure was the same with that described in Section 3.2. We checked the catalytic activity and DNA-binding affinity of the home-expressed Cas9 and confirmed that they were comparable to those of

the purchased Cas9 upon our experimental conditions.

Guide RNAs were prepared by *in vitro* transcription as described in Section 3.2 or purchased from ST Pharm, while all Cy5-labeled crRNAs were purchased from ST Pharm with HPLC-purification. Just as for experiments in Chapter 3, all RNA-including solutions were prepared with nuclease-free water with all disposable tips and microcentrifuge tubes certified as RNase-free and all experimental devices and lab space pre-treated with 70% ethanol.

All DNA oligomers with 5'-end biotin and/or internal amino modification were purchased from Integrated DNA Technologies, and dye-labeled double-stranded DNAs were prepared as already described in Section 3.2.

All the RNA and DNA sequences with the positions of biotin and internal amino modifications are listed in Appendix.

Single-molecule assays: The same single-molecule flow chamber and surface-passivation process were employed in this study with those described in Section 3.2. All imaging and binding/cleavage reactions were performed with the following buffer composition: 100 mM NaCl, 50 mM Tris-HCl pH 7.9, 10 mM MgCl2, 1 mM DTT and 0.1 mg/ml BSA.

In cleavage/binding measurements and smFRET experiments for FRET histograms, the imaging and binding/cleavage reactions were performed at 37 °C with oxygen scavenger (2.7 U/ml of pyranose oxidase (Sigma-Aldrich), 7.5 U/ml of catalase (Sigma-Aldrich) and 0.4% (w/v) of D-glucose) and triplet quencher (~2 mM Trolox) as described in Section 3.2. The DNA-cleavage efficiencies were measured

after 5-min incubation of immobilized DNA with 2 nM Cas9 and 300 nM gRNAs, and the binding affinities and FRET histograms were obtained after 30-min incubation of the DNA with 2 nM Cas9 and 30 nM gRNAs with the temporal resolution of 500 ms.

In smFRET experiments for single-molecule time traces, imaging was performed at room temperature with different oxygen scavenger (1 mg/ml of glucose oxidase (Sigma-Aldrich), 0.04 mg/ml of catalase (Sigma-Aldrich) and 0.8% (w/v) of D-glucose) and the addition of 5% (v/v) of glycerol. The time traces were acquired intermittently during the incubation (from 0 min to 30 min) of the DNA with Cas9:gRNA with the temporal resolution of 30 ms.

Set-up and data analysis: The single-molecule fluorescence assays employed our home-built total internal reflection fluorescence microscopy which is described in Chapter 2. The only difference is that for smFRET assays, we used a 532 nm green laser (Samba, Cobolt AB) alone to excite Cy3 molecules, whereas for cleavage assays, a 633 nm He-Ne laser (25-LHP-925-230, Melles-Griot) along with the green laser were both turned on to excite Cy3 and Cy5 molecules independently.

The procedure for the single-molecule cleavage assay was based on the quantification of the number of Cy3 and Cy5 molecules, which was identical to that described in Section 3.2. The same methodology was also applied to determine the bound ratio of Cy5-labeled Cas9:gRNA to Cy3-labeled DNA.

In the smFRET histograms and time traces, the FRET efficiency was calculated according to the following equation: $\beta I_{acceptor}/(\gamma I_{donor} + \beta I_{acceptor})$. The

gamma factor, γ, is the correction factor, determined as the ratio of change in the acceptor intensity, $\Delta I_{acceptor}$, to change in the donor intensity, ΔI_{donor}, when the acceptor photo-bleaches in the absence of protein: $\gamma = \Delta I_{acceptor}/\Delta I_{donor} = 1$ for the Cy3-Cy5 pair in our experimental conditions. The other correction factor β was introduced reflect the enhancement of acceptor fluorescence intensity upon protein binding: the value of β was determined to be 0.5 in our system. To identify FRET states and to calculate the transition rates between the states from the single-molecule time traces, hidden Markov modelling (HMM) based on variational Bayesian expectation maximization was employed[32]. On the other hand, to determine high-FRET-state populational ratio from the FRET histogram, an appropriate threshold value of FRET efficiency was set to sort high- and low-FRET states.

4.3. Results and Discussion

4.3.1. Real-time observation of Cas9:gRNA binding to DNA at single-molecule level

For Cas9-induced DNA-cleavage reaction, Cas9:gRNA should bind to the target DNA, where unidirectional R-loop expansion between the DNA and crRNA seems to be critical for the stabilization and nuclease-activation of Cas9:gRNA:DNA (Section 4.1). To search dynamic aspect of the R-loop structure and thereby, to elucidate molecular behavior of crRNA during the interaction of Cas9:gRNA with DNA, we used single-molecule fluorescence resonance energy transfer (smFRET) spectroscopy. A FRET pair between immobilized target DNA and crRNA was

introduced to monitor the conformational dynamics during the DNA-crRNA hybridization (Fig. 4.1a); Cy3 (the donor) labeled at the 25th nucleotide from the PAM site (+25) in non-complementary strand of the DNA and Cy5 (acceptor) at the 5´-end (20th nucleotide from the PAM site: +20) of crRNA. The real-time smFRET observation was successfully demonstrated with the Cy5-Cas9:gRNA incubated with 'wild-type' Cy3-DNA which contains a protospacer with full complementarity to crRNA and the right PAM sequence (5'-NGG-3'). After the injection of the mixture of Cas9, tracrRNA, and Cy5-crRNA onto the Cy3-DNA, Cy5 molecules, fluoresced by FRET from excited Cy3 molecules, appeared gradually, each spot of which represented individual real-time binding of Cas9:gRNA to DNA (Fig. 4.1b and c).

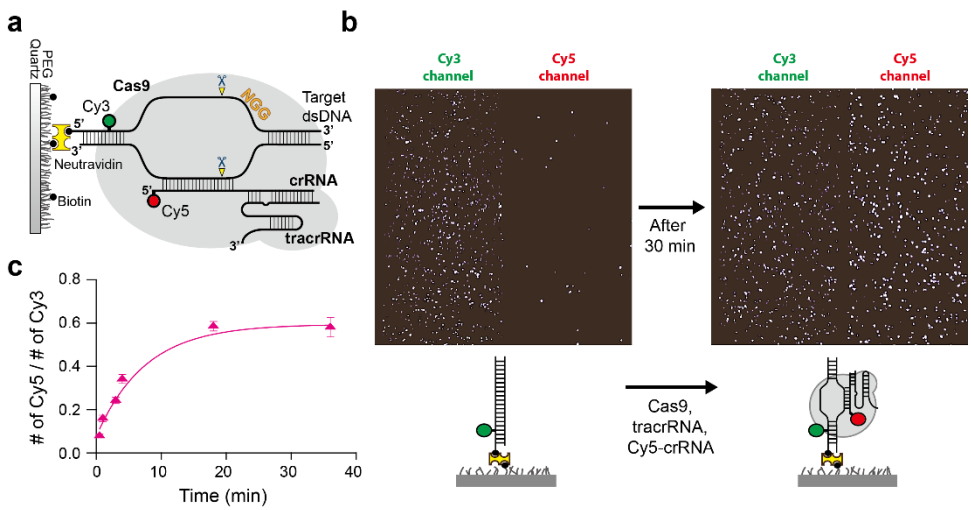

Figure 4.1. Real-time observation of target DNA-Cas9:gRNA binding using single-molecule FRET. (a) Cartoon for surface-immobilized Cy3-labeled target DNA, bound to Cy5-labeled Cas9:gRNA. (b) CCD images showing single-molecules of the DNA and the bound Cas9 complex before and after the incubation of Cas9:gRNA with DNA. (c) Kinetic curve for Cas9:gRNA binding to the target DNA.

4.3.2. Conformational heterogeneity in crRNA-DNA heteroduplex

Using the smFRET technique, we observed the structural dynamics in the crRNA-DNA heteroduplex. After Cas9:gRNA and the wild-type DNA equilibrated by sufficient incubation (30 min), a FRET histogram exhibited two distinct states of Cas9:gRNA:DNA: a high-FRET state (FRET value of 0.83) and a low-FRET state (FRET value of 0.27) (Fig. 4.2a, upper histogram). Also, for dCas9 (catalytically 'dead' Cas9, point-mutated at the catalytically active sites, D10A and H840A) in place of Cas9, the identical FRET states were observed (Fig. 4.2a, lower histogram), indicating that these two different FRET states may represent the sub-conformational intermediates during the DNA binding and cleavage process of Cas9:gRNA.

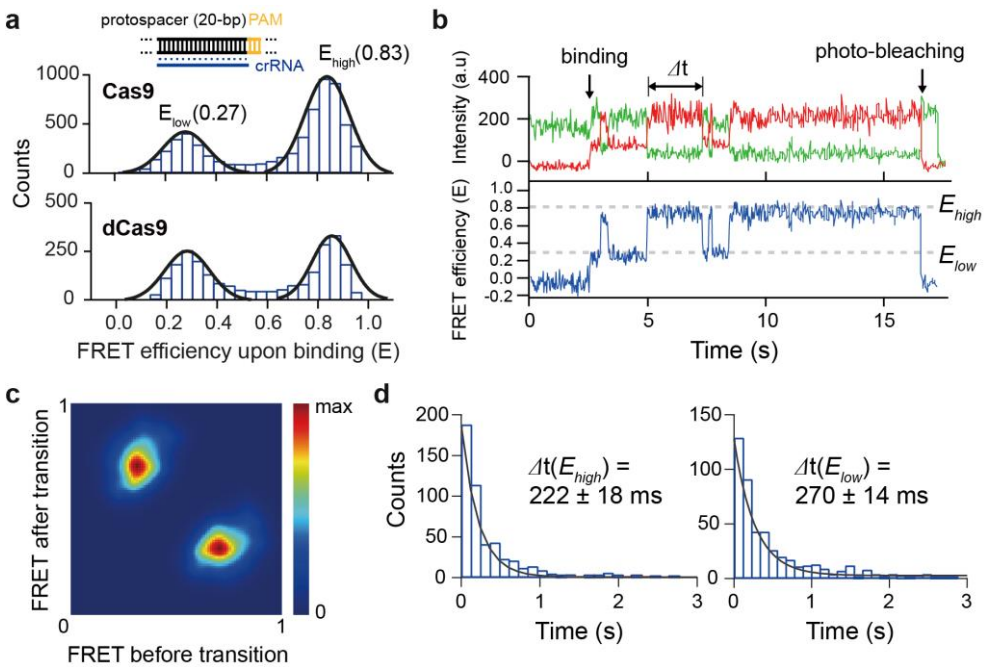

Figure 4.2. Conformational heterogeneity and dynamics in Cas9:gRNA:DNA. (a) FRET histograms for the crRNA-DNA heteroduplex with Cas9 and dCas9. (b) A representative single-molecule time-trace for Cas9:gRNA:DNA. (c) Transition density plot for all transitions in the time traces. (d) Dwell-time analysis for the high and low-FRET states.

To characterize the detailed conformation of each FRET state, we looked into single-molecule time traces and designed two other control DNA samples. In the single-molecule time traces, the nascent complexes of Cas9:gRNA initially bound to the wild-type DNA mostly presented the low-FRET value (Fig. 4.2b). Furthermore, when the complementary strand of DNA (complementary to the crRNA) and crRNA on the PAM-distal region (+15 ~ +20) are designed to be mismatched (termed 'mis-crRNA'), only a single low-FRET peak at the value of 0.27 was observed (Fig. 4.3, upper panel). These results imply that the low-FRET state corresponds to the PAM-proximal bound but PAM-distal unbound complex (termed 'open conformation'), in which crRNA hybridizes with the target DNA only partially on the PAM-proximal seed region before the completion of the R-loop expansion.

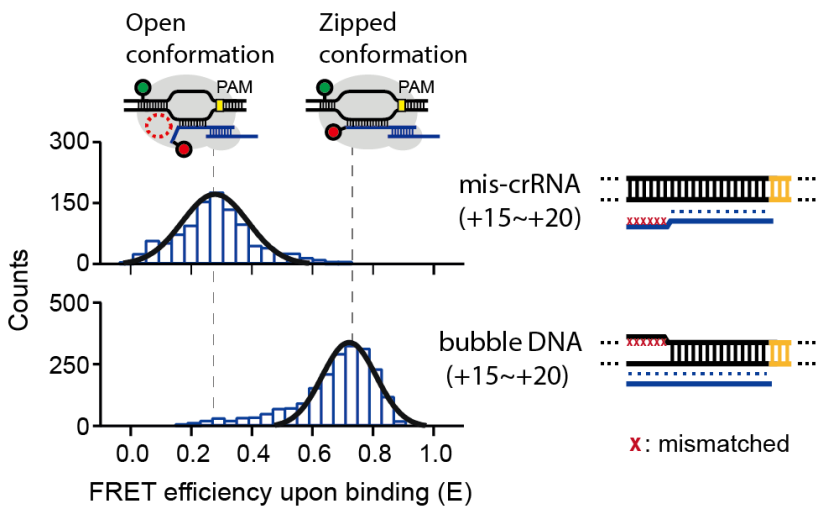

Figure 4.3. Characterization of the sub-conformations using mismatched DNAs. FRET histograms for the two DNAs modified on PAM-distal region (+15 ~ +20). 'Mis-crRNA' refers to the intact dsDNA, in which the complementary strand is mismatched to crRNA on the PAM-distal region, whereas 'bubble DNA' refers to the bulged dsDNA, in which the complementary and non-complementary strands are mismatched to each other on the PAM-distal region with the complementary strand fully matched to crRNA.

On the other hand, following the formation of the open conformation of Cas9:gRNA with the wild-type DNA, transition toward the high-FRET state was observed (Fig. 4.2b). Taking available X-ray crystal structures of Cas9:gRNA:DNA into account[16,17], the high-FRET state was thought to represent the conformation after the R-loop completion (termed 'zipped conformation') with the observed FRET transition from the low-FRET to the high-FRET value relevant to the R-loop expansion on the PAM-distal region. Plus, the FRET value of the high-FRET state (0.83), calculated to be inter-dye distance of 40 Å, was compatible with the zipped R-loop structure considering the closeness between the 5′-end of crRNA and the +25 position of the non-complementary strand of DNA, revealed by the recent electron microscopic study[20].

Surprisingly, the R-loop structure did not remain in the zipped conformation even after it transitioned from the open conformation. Rather, a series of transient, repetitive transitions between the two conformations was detected (Fig. 4.2b). A transition density plot from all the transitions in the single-molecule time traces showed only the two peaks (from the low-FRET state to the high-FRET state, and vice versa) (Fig. 4.2c), indicating that only the open and zipped conformations are the well-defined states in the intrinsically dynamic R-loop structure. In terms of kinetics, the dwell-time distributions of the two states were both fitted well by single exponential functions ($\Delta t(E_{high})$ = 222 ms ± 18 ms and $\Delta t(E_{low})$ = 270 ms ± 14 ms), confirming that each conformation is kinetically a single state (Fig. 4.2d). Additionally, removing the complementarity between the two DNA strands on the PAM-distal region (+15 ~ +20) (termed 'bubble DNA' design), only the high-FRET

state was observed in the FRET histogram (Fig. 4.3, bottom panel) without any FRET transitions (we note that the high-FRET value of bubble DNA (0.75) is evidently lower than that of wild-type DNA (0.83), probably due to the structural distortion of the bubble DNA induced by the mismatched bases). This observation reinforces our identification that the repetitive FRET transition represents the conformational dynamics between crRNA and the complementary strand of DNA at the PAM-distal end and that the high-FRET state represents the fully expanded R-loop conformation, the zipped conformation. Together, these results clearly demonstrated the sub-steps of the R-loop expansion in Cas9:gRNA:DNA with the direct observation of the conformational sub-populations in the R-loop structure.

4.3.3. crRNA-controlled conformational distribution in R-loop structure in Cas9:gRNA:DNA

Next, to identify the role of crRNA in the PAM-distal structural dynamics in Cas9:gRNA:DNA, truncated crRNA (17-nt crRNA) was employed, in which three nucleotides were deleted at the PAM-distal end. We hypothesized that the loss of the interaction between truncated crRNA and the complementary strand of DNA on the PAM-distal region may lead to a shift in the conformational equilibrium between the two R-loop sub-conformations. In agreement with the hypothesis, the FRET histogram with a considerably altered distribution was measured after sufficient incubation (30 min) with truncated crRNA. In comparison to full-length crRNA (20-nt crRNA), also the two FRET states were observed whereas their relative

populations were reversed (Fig. 4.4a), indicating that the truncation shifts the equilibrium toward base-pairing between the two DNA strands on the PAM-distal region, i.e. the open conformation (notably, the FRET values of the low and high FRET states changed slightly (0.29 and 0.76, respectively), possibly due to the change in the inter-dye distance by the truncation of the 3-nt in each state). Kinetically, the average dwell time of the high-FRET state (77 ms ± 9 ms) was analyzed to become 3 times shorter than that of 20-nt crRNA (222 ms ± 18 ms), while the dwell time of the low-FRET state remained nearly unchanged (250 ms ± 12 ms for 17-nt crRNA and 270 ms ± 14 ms for 20-nt crRNA) (Fig. 4.4b and c).

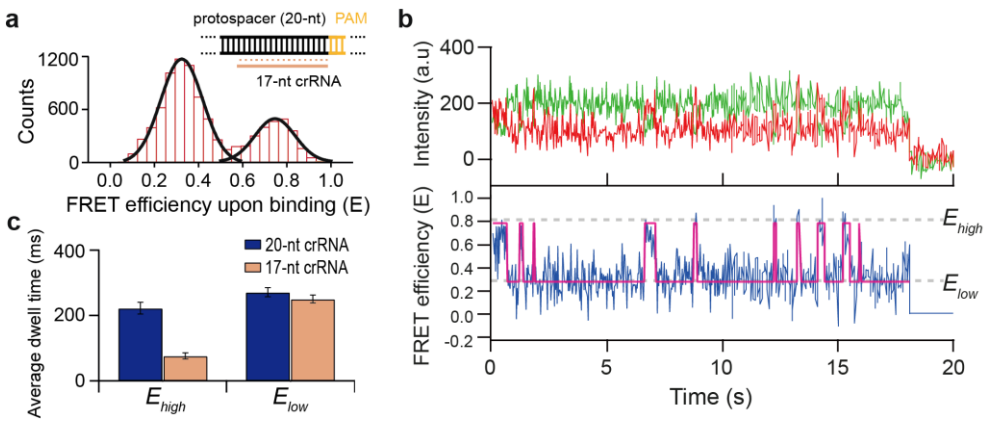

Figure 4.4. Effect of PAM-distal complementarity on the sub-conformational distribution and kinetics. (a) FRET histogram for Cas9:gRNA with 17-nt crRNA bound to the wild-type DNA. (b) A representative single-molecule time-trace for Cas9:gRNA:DNA with 17-nt crRNA. (c) Comparision of each dwell time of the high and low-FRET states between 20-nt full-length and 17-nt truncated crRNA (mean ± SEM, n ≥ 3).

These data support that the nucleotide shortening causes a lower binding energy of the PAM-distal crRNA-DNA heteroduplex than that with 20-nt crRNA, thereby destabilizing the zipped conformation without disturbing the stability of the

open conformation. This is also in good agreement with the previous report proposing that the underlying principle of improved target specificity of truncated crRNA is the 3-nt deletion-derived destabilization of the RNA-DNA interaction to a level just endurable for on-target activity but unendurable for off-target activity[27]. In other words, our results demonstrated that the PAM-distal complementarity of crRNA to the complementary strand of DNA kinetically affects the conformational transition and thereby the distribution between the two R-loop sub-conformations.

4.3.4. Nuclease activity of Cas9 regulated by the conformational transition

With the aim of correlating the R-loop sub-structural transition with the nuclease activity of Cas9, we compared the DNA-cleavage efficiencies of the three DNA designs (wild-type DNA, bubble DNA, and mis-crRNA), all of which exhibited the different conformational distributions between the open and zipped conformations (Fig. 4.2a and 4.3). The wild-type and bubble DNA, in which the zipped conformations were available, were cleaved by the Cas9 complex with comparable efficiencies, whereas the mis-crRNA, in which binding affinity was at about the same level with the other DNAs but remained only in the open conformation, failed to be cleaved (Fig. 4.5a). Moreover, kinetic comparison for the Cas9-catalyzed DNA-cleavage reaction of the wild-type (cleavage rate constant of 0.58/min) and bubble DNA (cleavage rate constant of 1.59/min) suggests that the shorter Cas9:gRNA:DNA spends time in the zipped conformation, the slower DNA cleavage

occurs (Fig. 4.5b, left plot). This interpretation was further corroborated by the kinetics of the DNA-cleavage reaction with 17-nt crRNA (Fig. 4.5b, right plot), in which the DNA-cleavage rate (0.07/min) was remarkably reduced due to the decrease in the dwell time of the zipped conformation (Fig. 4.4c). These results reveal that once the Cas9 complex binds to DNA, dsDNA melting with the crRNA-DNA heteroduplexation on the PAM-distal region is the rate-determining step for the DNA-cleavage reaction; thus, the DNA cleavage is allowed only when the R-loop remains stably in the zipped conformation.

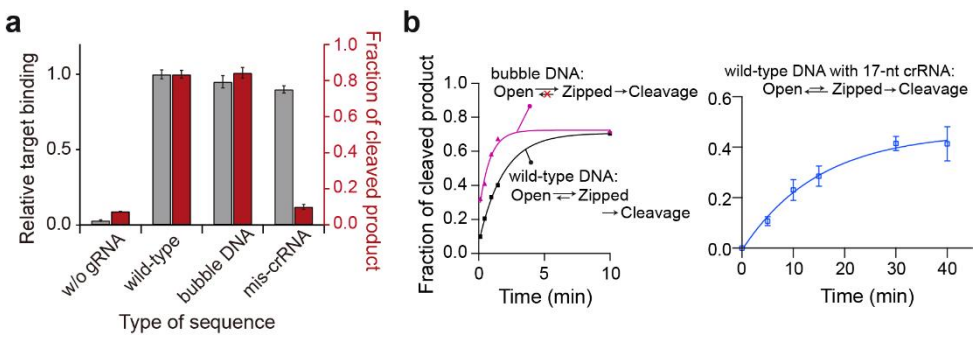

Figure 4.5. Nuclease activation of Cas9:gRNA by the sub-conformational transition. (a) Relative binding affinities and DNA-cleavage activities of Cas9:gRNA for the three afore-described DNA constructs (mean ± SEM, n ≥ 5). (b) Kinetic curves of the cleaved DNA ratio for the wild-type, bubble DNA and the wild-type DNA with 17-nt crRNA (error bars represent s.d.).

Finally, we varied the R-loop stability using a series of off-target DNAs with 2-3 mismatched bases from the PAM-proximal to PAM-distal end for the comprehensive picture of the relation between the stability of the R-loop structure and the DNA-cleavage efficiency of the Cas9 complex from the PAM-proximal to PAM-distal region. Cas9:gRNA failed to form the stable Cas9:gRNA:DNA complex

with the DNAs mutated in the PAM-proximal 'seed' region (Fig. 4.6a), as previously reported[25]; the binding affinity of Cas9:gRNA to DNAs arose with at least half of the protospacer sequences from the PAM site matched to crRNA. As the mismatched position moving toward the PAM-distal end, the DNA-cleavage efficiency increased along with the populational increase in the zipped conformation (Fig. 4.6b), consistent with the enhancement and reduction of the cleavage efficiency for the bubble DNA and 17-nt crRNA construct, respectively. To note, the DNA containing mismatches at +10 ~ +11 positions did not transition to the high-FRET state, despite 9-nt complementarity at the PAM-distal end. This implies that the zipped conformation requires not only the crRNA-DNA complementarity on the PAM-distal region but the entire heteroduplex stability including the PAM-proximal region, which is consistent with the previously proposed unidirectionality of the R-loop expansion. Overall, these findings manifest that the observed transition between the two conformations is implicated in the enzymatic activation of the Cas9 complex as well as the R-loop expansion, rendering the conformational rearrangement a key step for the nuclease activation in the Cas9:gRNA:DNA structure.

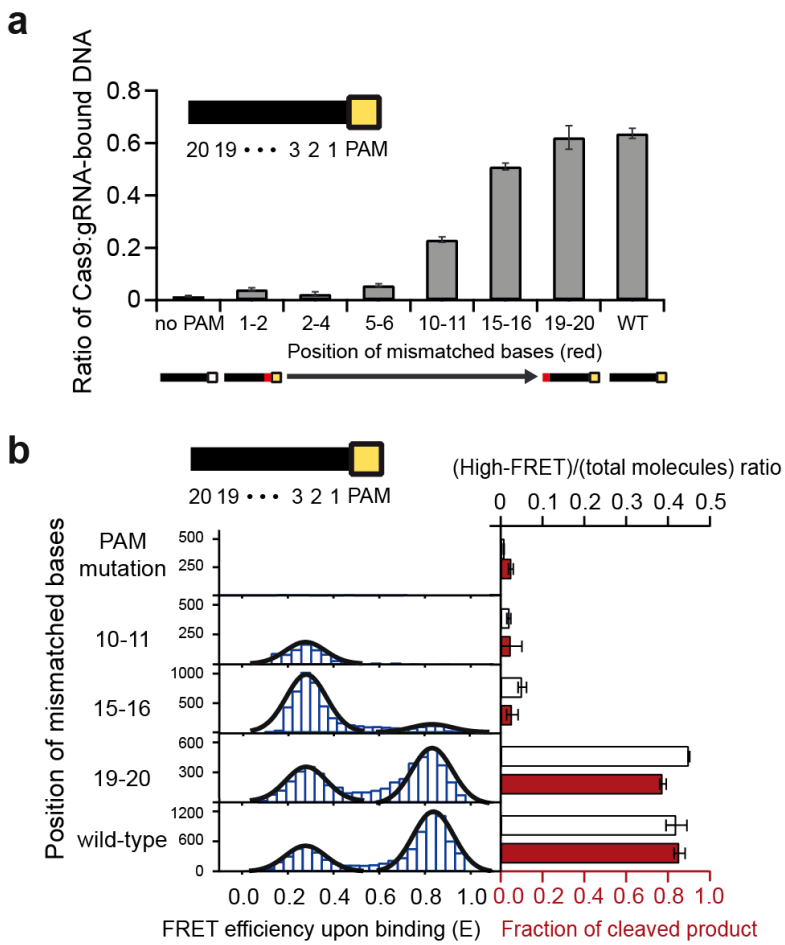

Figure 4.6. Binding affinity, sub-conformational distribution, and DNA-cleavage activity with various mismatched DNAs. (a) Binding affinities of Cas9:gRNA to various off-target DNAs, in which the crRNA-complementarity is removed at different positions. (b) FRET histograms of the off-target DNAs and the comparison of the high-FRET-state ratio with the cleaved-DNA ratio (mean ± SEM, n ≥ 5).

4.4. Conclusion

To summarize, this chapter reported our real-time single-molecule experiments on the structural dynamics of crRNA-target DNA heteroduplex, which was found to be regulated by the heteroduplex stability and to regulate the Cas9 nuclease activation. The R-loop expansion in Cas9:gRNA:DNA was revealed to be a multistep process with the dynamic transitions on the PAM-distal region between the two intermediate states, open and zipped conformations (Fig. 4.7). This intrinsic structural dynamics is consistent with the previously suggested structural flexibility of the Cas9 complex deduced from the X-ray crystallography[17,18] and with a previous biophysical report in which Cas9, in contrast to type-I CRISPR effector complex, was found not to be locked after the R-loop completion[26].

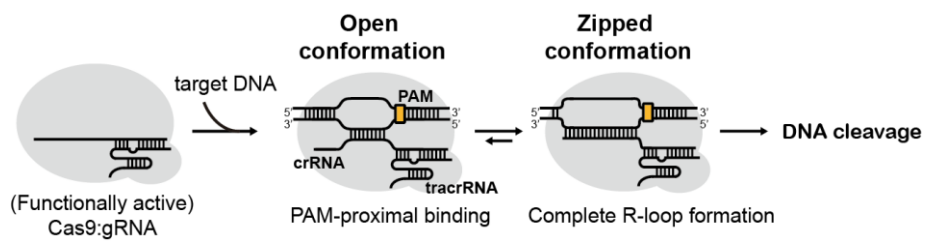

Figure 4.7. Schematic model for the multistep R-loop expansion, followed by DNA-cleavage reaction.

The transition rate between the two conformations (Fig. 4.2d) was two orders of magnitude faster than the DNA-cleavage rate of the wild-type DNA (Fig. 4.5b). This discordance of the two rate constants indicates that the formation of the zipped conformation alone is insufficient for DNA cleavage although the transition toward the zipped conformation is the prerequisite for the cleavage reaction. As several structure-based studies have proposed that the HNH nuclease domain of Cas9 moves

from the PAM-distal region to PAM-proximal cleavage site of the complementary strand of DNA[17,18,28], it is possible that the transition to the zipped conformation may trigger the nuclease domain movement; synchronous real-time single-molecule observation for the crRNA-DNA heteroduplex with the HNH domain is required to elucidate the correspondence between the two structural rearrangements.

Off-target DNA sequences with mismatched bases on the PAM-distal region formed stable Cas9:gRNA:DNA complex but remained only in the PAM-proximally bound, open conformation without transition to the zipped conformation and the following DNA cleavage (Fig. 4.5 and 4.6). These experiments with the variable off-target DNA sequences revealed that the stability of the crRNA-DNA heteroduplex controlled the sub-conformational kinetics and distribution in Cas9:gRNA:DNA, thereby regulating the nuclease activation of Cas9. In light of this newly unveiled role of the kinetics and dynamics in Cas9:gRNA:DNA, it would become possible to predict off-target activities more quantitatively, which enables more specific target sequence design for genome-wide gene editing.

5. Conclusion

In this thesis, we have discussed our single-molecule and (single-molecule-based) pseudo-ensemble studies on the molecular mechanism of CRISPR/Cas9 functioning, focusing on the structural roles of gRNAs. To summarize, the background principles and characteristics of the single-molecule fluorescence spectroscopic technique and CRISPR/Cas system employed in our studies were firstly described. Then, the distinct structural roles of tracrRNA and crRNA identified in our studies were illustrated sequentially with the details of material preparation and experimental procedures.

With the single-molecule-based pseudo-ensemble DNA cleavage assay for different sets of partial pre-incubation processes, conformational change from a normally active *apo*-Cas9 state to an inactivated *apo*-Cas9 state in the absence of tracrRNA was newly uncovered. The transition toward inactivated *apo*-Cas9 was blocked by tracrRNA-Cas9 interaction, leading to the formation of functional Cas9:gRNA. On the other hand, real-time monitoring of Cas9:gRNA:DNA using the single-molecule FRET assay characterized sub-conformations (open and zipped conformation) of the R-loop structure between crRNA and target DNA in Cas9:gRNA:DNA, which could not be observable from any other ensemble-averaged techniques. The importance of this transition between the sub-conformations was emphasized by its regulatory effect on the nuclease activity of Cas9:gRNA:DNA.

Owing to the powerfulness of the single-molecule spectroscopic technique, we distinguished the hidden intermediate states (i.e. inactivated *apo*-Cas9 and open and zipped sub-R-loop conformations) and analyzed their roles in the reaction

pathway of CRISRP/Cas9 functioning. Undoubtedly, there still remain uncertain points to be addressed: structural characterization of inactivated *apo*-Cas9, a kinetic model for the tracrRNA-involved recovery of the nuclease activity from the inactivated state, discrimination of the sub-conformational distributions before and after the DNA cleavage, and synchronousness between the open-to-zipped conformational transition and the HNH-nuclease-domain movement. Nonetheless, we expect that our study would serve as a doorstep to the complete understanding of the mechanism of action of the CRISPR/Cas9 system.

References

1. Jansen, R., van Embden, J. D. A., Gaastra, W. & Schouls, L. M. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **43,** 1565–1575 (2002).

2. Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151,** 2551–2561 (2005).

3. Marraffini, L. A. & Sontheimer, E. J. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat. Rev. Genet.* **11,** 181–190 (2010).

4. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315,** 1709–1712 (2007).

5. Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471,** 602–607 (2011).

6. Kim, H. & Kim, J.-S. A guide to genome engineering with programmable nucleases. *Nat. Rev. Genet.* **15,** 321–334 (2014).

7. Makarova, K. S. *et al.* Evolution and classification of the CRISPR–Cas systems. *Nat. Rev. Microbiol.* **9,** 467–477 (2011).

8. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155,** 733–740 (2009).

9. Deveau, H. *et al.* Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *J. Bacteriol.* **190,** 1390–1400 (2008).

10. Jinek, M. *et al.* A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science* **337,** 816–821 (2012).

11. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339,** 819–823 (2013).

12. Jinek, M. *et al.* RNA-programmed genome editing in human cells. *Elife* **2,** e00471 (2013).

13. Cho, S. W., Kim, S., Kim, J. M. & Kim, J.-S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* **31,** 230–232 (2013).

14. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339,** 823–826 (2013).

15. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA

interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507,** 62–67 (2014).

16. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513,** 569–573 (2014).

17. Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156,** 935–949 (2014).

18. Jiang, F., Zhou, K., Ma, L., Gressel, S. & Doudna, J. A. Structural biology. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science* **348,** 1477–1481 (2015).

19. Jinek, M. *et al.* Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343,** 1247997 (2014).

20. Jiang, F. *et al.* Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* **351,** 867–871 (2016).

21. Förster, T. Zwischenmolekulare energiewanderung und fluoreszenz. *Ann. Phys.* **2,** 55–75 (1948). An English version is available on Mielczarek, E. V., Greenbaum, E. & Knox, R. S. *Biological Physics*. American Institute of Physics:New York, 148-160 (1993).

22. Benesch, R. E. & Benesch, R. Enzymatic removal of oxygen for polarography and related methods. *Science* **118,** 447–448 (1953).

23. Swoboda, M. *et al.* Enzymatic oxygen scavenging for photostability without pH drop in single-molecule experiments. *ACS Nano* **6,** 6364–6369 (2012).

24. Cordes, T., Vogelsang, J. & Tinnefeld, P. On the mechanism of Trolox as antiblinking and antibleaching reagent. *J. Am. Chem. Soc.* **131,** 5018–5019 (2009).

25. Wu, X. *et al.* Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat. Biotechnol.* **32,** 670–676 (2014).

26. Szczelkun, M. D. *et al.* Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc. Natl. Acad. Sci. USA* **111,** 9798–9803 (2014).

27. Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. & Joung, J. K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.* **32,** 279–284 (2014).

28. Sternberg, S. H., LaFrance, B., Kaplan, M. & Doudna, J. A. Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature* **527,** 110–113 (2015).

29. Selvin, P. R. & Ha, T. *Single-molecule techniques: a laboratory manual*.

Cold Springer Harbor Laboratory Press: New York, 18-19 (2008).

30. Roy, R., Hohng, S. & Ha, T. A practical guide to single-molecule FRET. *Nat. Methods* **5,** 507–516 (2008).

31. Rutkauskas, M. *et al*. Directional R-loop formation by the CRISPR-cas surveillance complex Cascade provides efficient off-target site rejection. *Cell Rep.* **10,** 1534–1543 (2015).

32. Bronson, J. E., Fei, J., Hofman, J. M., Gonzalez, R. L. J. & Wiggins, C. H. Learning rates and states from biophysical time series: a Bayesian approach to model selection and single-molecule FRET data. *Biophys. J.* **97,** 3196–3205 (2009).

Appendix

All DNA and RNA sequences, used in Chapter 3 and 4, are listed below with the introduced modifications:

Description	Type		Sequences
wild-type DNA in cleavage assay (Ch. 3 and 4)	Non-complementary strand		5’ – /5Biosg/ttt ttt GAG GAA G**/iAmMC6T/**G CCT GAG TCC GAG CAG AAG AAG AAG GGC TCC CAT CAC ATC – 3’
	Complementary strand		5’ – GAT GTG A**/iAmMC6T/**G GGA GCC CTT CTT CTT CTG CTC GGA CTC AGG CAC TTC CTC – 3’
wild-type DNA in smFRET assay (Ch. 4)	Non-complementary strand		5’ – /5Biosg/ttt ttt GAG GAA G**/iAmMC6T/**G CCT GAG TCC GAG CAG AAG AAG AAG GGC TCC CAT CAC ATC – 3’
	Complementary strand		5’ – GAT GTG ATG GGA GCC CTT CTT CTT CTG CTC GGA CTC AGG CAC TTC CTC – 3’
mis-crRNA (Ch. 4)	Non-complementary strand		5’ – /5Biosg/ttt ttt GAG GAA G**/iAmMC6T/**G CCT **CTC AGG** GAG CAG AAG AAG AAG GGC TCC CAT CAC ATC – 3’
	Complementary strand	For smFRET assay	5’ – GAT GTG ATG GGA GCC CTT CTT CTT CTG CTC **CCT GAG** AGG CAC TTC CTC – 3’
		For cleavage assay	5’ – GAT GTG A**/iAmMC6T/**G GGA GCC CTT CTT CTT CTG CTC **CCT GAG** AGG CAC TTC CTC – 3’
bubble DNA (Ch. 4)	Non-complementary strand		5’ – /5Biosg/ttt ttt GAG GAA G**/iAmMC6T/**G CCT **CTC AGG** GAG CAG AAG AAG AAG GGC TCC CAT CAC ATC – 3’
	Complementary strand	For smFRET assay	5’ – GAT GTG ATG GGA GCC CTT CTT CTT CTG CTC GGA CTC AGG CACTTC CTC – 3’
		For cleavage	5’ – GAT GTG A**/iAmMC6T/**G GGA GCC CTT CTT CTT CTG

		assay	CTC GGA CTC AGG CAC TTC CTC – 3’
no-PAM off-target DNA (Ch. 4)	Non-complementary strand		5’ – /5Biosg/ttt ttt GAG GAA G**/iAmMC6T/**G CCT GAG TCC GAG CAG AAG AAG AA**C CC**C TCC CAT CAC ATC – 3’
	Comple-mentary strand	For smFRET assay	5’ – GAT GTG ATG GGA G**GG G**TT CTT CTT CTG CTC GGA CTC AGG CAC TTC CTC – 3’
		For cleavage assay	5’ – GAT GTG A**/iAmMC6T/**G GGA G**GG G**TT CTT CTT CTG CTC GGA CTC AGG CAC TTC CTC – 3’
1-2 mismatched off-target DNA (Ch. 4)	Non-complementary strand		5’ – /5Biosg/ttt ttt GAG GAA G**/iAmMC6T/**G CCT GAG TCC GAG CAG AAG AAG **TT**G GGC TCC CAT CAC ATC – 3’
	Comple-mentary strand	For smFRET assay	5’ – GAT GTG ATG GGA GCC C**AA** CTT CTT CTG CTC GGA CTC AGG CAC TTC CTC – 3’
		For cleavage assay	5’ – GAT GTG A**/iAmMC6T/**G GGA GCC C**AA** CTT CTT CTG CTC GGA CTC AGG CAC TTC CTC – 3’
2-4 mismatched off-target DNA (Ch. 4)	Non-complementary strand		5’ – /5Biosg/ttt ttt GAG GAA G**/iAmMC6T/**G CCT GAG TCC GAG CAG AAG A**TC T**AG GGC TCC CAT CAC ATC – 3’
	Comple-mentary strand	For smFRET assay	5’ – GAT GTG ATG GGA GCC CT**A GA**T CTT CTG CTC GGA CTC AGG CAC TTC CTC – 3’
		For cleavage assay	5’ – GAT GTG A**/iAmMC6T/**G GGA GCC CT**A GA**T CTT CTG CTC GGA CTC AGG CAC TTC CTC – 3’
5-6 mismatched off-target DNA (Ch. 4)	Non-complementary strand		5’ – /5Biosg/ttt ttt GAG GAA G**/iAmMC6T/**G CCT GAG TCC GAG CAG AA**C T**AG AAG GGC TCC CAT CAC ATC – 3’
	Comple-mentary	For smFRET assay	5’ – GAT GTG ATG GGA GCC CTT CT**A G**TT CTG CTC GGA CTC AGG CAC TTC CTC – 3’

	strand	For cleavage assay	5’ – GAT GTG A/**iAmMC6T**/G GGA GCC CTT CT**A G**TT CTG CTC GGA CTC AGG CAC TTC CTC – 3’
10-11 mismatched off-target DNA (Ch. 4)	Non-complementary strand		5’ – /5Biosg/ttt ttt GAG GAA G/**iAmMC6T**/G CCT GAG TCC GAG **GT**G AAG AAG AAG GGC TCC CAT CAC ATC – 3’
	Comple-mentary strand	For smFRET assay	5’ – GAT GTG ATG GGA GCC CTT CTT CTT C**AC** CTC GGA CTC AGG CAC TTC CTC – 3’
		For cleavage assay	5’ – GAT GTG A/**iAmMC6T**/G GGA GCC CTT CTT CTT C**AC** CTC GGA CTC AGG CAC TTC CTC – 3’
15-16 mismatched off-target DNA (Ch. 4)	Non-complementary strand		5’ – /5Biosg/ttt ttt GAG GAA G/**iAmMC6T**/G CCT GAG T**TA** GAG CAG AAG AAG AAG GGC TCC CAT CAC ATC – 3’
	Comple-mentary strand	For smFRET assay	5’ – GAT GTG ATG GGA GCC CTT CTT CTT CTG CTC **TA**A CTC AGG CAC TTC CTC – 3’
		For cleavage assay	5’ – GAT GTG A/**iAmMC6T**/G GGA GCC CTT CTT CTT CTG CTC **TA**A CTC AGG CAC TTC CTC – 3’
19-20 mismatched off-target DNA (Ch. 4)	Non-complementary strand		5’ – /5Biosg/ttt ttt GAG GAA G/**iAmMC6T**/G CCT **CT**G TCC GAG CAG AAG AAG AAG GGC TCC CAT CAC ATC – 3’
	Comple-mentary strand	For smFRET assay	5’ – GAT GTG ATG GGA GCC CTT CTT CTT CTG CTC GGA C**AG** AGG CAC TTC CTC – 3’
		For cleavage assay	5’ – GAT GTG A/**iAmMC6T**/G GGA GCC CTT CTT CTT CTG CTC GGA C**AG** AGG CAC TTC CTC – 3’
crRNA (Ch. 3 and 4)	ssRNA for cleavage assay		5’ – GGG AGU CCG AGC AGA AGA AGA AGU UUU AGA GCU AUG CUG UUU UG – 3’
tracrRNA (Ch. 3 and 4)	ssRNA for both cleavage and		5’–GGA ACC AUU CAA AAC AGC AUA GCA AGU UAA AAU AAG GCU AGU CCG UUA

	smFRET assay	UCA ACU UGA AAA AGU GGC ACC GAG UCG GUG CUU UUU U– 3’
Cy5-crRNA (Ch. 4)	ssRNA for smFRET assay	5’ – /5Cy5/ GAG UCC GAG CAG AAG AAG AAG UUU UAG AGC UAU GCU GUU UUG – 3’
17-nt Cy5-crRNA (Ch. 4)	ssRNA for smFRET assay	5’ – /5Cy5/ UCC GAG CAG AAG AAG AAG UUU UAG AGC UAU GCU GUU UUG – 3’
17-nt crRNA (Ch. 4)	ssRNA for cleavage assay	5’ – UCC GAG CAG AAG AAG AAG UUU UAG AGC UAU GCU GUU UUG – 3’

“Biosg” refers to biotin for surface immobilization, and “iAmMC6T” refers to amino-modified carbon-spacer-linked dT for NHS-ester linked Cy3 or Cy5 dye labeling. For DNA constructs other than the wild type, the mutated sequences are underlined in bold.

국문초록

원핵생물의 후천면역체계로서 최근들어 새롭게 발견된 크리스퍼 시스템 (Clustered regularly interspaced short palindromic repeats (CRISPR) / CRISPR-associated (Cas) system) 은 외부의 바이러스나 플라스미드로부터 개체 스스로를 보호하는 기능을 수행한다. 여러 타입의 크리스퍼 시스템 중, 타입-II 시스템은 Cas9이라는 핵산절단효소 단백질과 두 가지의 가이드 RNA (gRNA: CRISPR RNA (crRNA) 와 trans-activating crRNA (tracrRNA))로 구성되며, 이 구성요소들이 단백질-RNA 복합체를 이루어 표적 DNA 를 인식하고 절단한다. 이 타입-II 크리스퍼/Cas9 시스템이 차세대 유전자 조작 기술로 유전공학 분야에 성공적으로 도입되면서, Cas9-gRNA 복합체에 대한 구조적 특성 연구도 활발히 진행되었다. 그러나 여전히 gRNA가 Cas9 핵산절단효소를 어떻게 활성화시키는지 그 구조변화에 대한 구체적인 동역학적 메커니즘에 대한 규명은 이루어지지 않은 실정이다.

이에, 우리는 단분자 형광 분광법을 도입하여 두 가이드 RNA가 Cas9 핵산절단효소 활성화에 있어 수행하는 구조적 역할에 대해 연구하였다. 그 결과, 정적인 앙상블 평균값만을 관찰하는 기존의 실험법들로는 알아낼 수 없었던 Cas9 시스템 작동 상의 세부적 메커니즘을 규명할 수 있었으며, 본 학위논문에서는 이 연구에 대해 논하기로 한다. 첫째로, tracrRNA는 정상 기능하는 Cas9-gRNA 복합체 (Cas9:gRNA)의 형성과정에

있어 핵심적인 역할을 수행함을 밝혀내었다. tracrRNA가 없는 경우, Cas9 단백질이 비활성화되어 Cas9:gRNA 형성이 이루어지지 않았다. 반면, tracrRNA-Cas9 상호작용이 가능한 경우에는 Cas9:gRNA가 형성되는 반응 경로가 우세해지면서 Cas9의 비활성화는 관찰되지 않았다. 둘째로, tracrRNA가 Cas9:gRNA 형성에 관여하는 것과는 달리, crRNA는 Cas9:gRNA가 표적 DNA와 결합한 Cas9-gRNA-DNA 삼중 복합체 (Cas9:gRNA:DNA)에서 Cas9 핵산절단효소의 활성화 여부를 조절하는 것으로 밝혀졌다. Cas9:gRNA:DNA 내의 crRNA와 DNA 사이에 형성되는 R-loop 구조는 '열린 구조 (open conformation)'와 '잠긴 구조 (zipped conformation)'라고 명명한 두 하위 구조로 존재하며, R-loop 구조는 실제적으로 이 두 구조를 계속해서 왔다갔다 반복하는 내재적인 동역학적 특성을 지니고 있음이 발견되었다. 나아가, 잠긴 구조가 만들어져야만 DNA가 절단된다는 것을 관찰함으로써, 이러한 crRNA와 연관된 구조 동역학적 특성이 Cas9 절단효소 활성화에 핵심임을 밝혀내었다. 종합하자면, 이러한 결과들은 Cas9 핵산절단효소의 구조적 활성화에 있어 gRNA의 역할에 대한 세부 메커니즘 수준에서의 이해를 가능케 하였다.

주요어: 단분자 분광학, 전반사 형광 현미경법, 형광 공명 에너지 전달, 크리스퍼, 구조 동역학, 효소 메커니즘

학 번: 2015-20386